MICROCOPY RESOLUTION TEST CHART
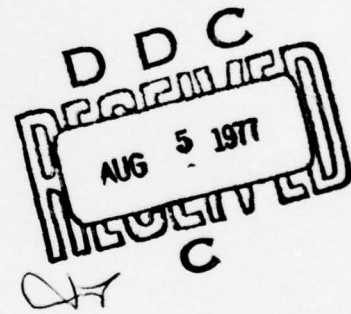NATIONAL BUREAU OF STANDARDS-1963-A

A STUDY OF PSYCHOACOUSTICS IN PASSIVE SONAR
CLASSIFICATION; PART 2, A REVIEW OF GENERAL CONCEPTS
AND A DISCUSSION OF RESULTS OBTAINED TO DATE (U)

C. P. Janota

Technical Memorandum
File No. TM 75-165
May 28, 1975
Contract No. N00017-73-C-1418

Copy No. _3_

The Pennsylvania State University
Institute for Science and Engineering
APPLIED RESEARCH LABORATORY
P. O. Box 30
State College, PA    16801

NAVY DEPARTMENT

NAVAL SEA SYSTEMS COMMAND

## DOCUMENT CONTROL DATA - R & D

*(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)*

| 1. ORIGINATING ACTIVITY *(Corporate author)* | 2a. REPORT SECURITY CLASSIFICATION |
|---|---|
| Applied Research Laboratory<br>P. O. Box 30<br>State College, PA 16801 | Unclassified |
| | 2b. GROUP |

3. REPORT TITLE

A Study of Psychoacoustics in Passive Sonar Classification. Part 2. A Review of General Concepts and A Discussion of Results Obtained to Date

4. DESCRIPTIVE NOTES *(Type of report and inclusive dates)*

5. AUTHOR(S) *(First name, middle initial, last name)*

Claus P. Janota

Technical memo.,

| 6. REPORT DATE | 7a. TOTAL NO. OF PAGES | 7b. NO. OF REFS |
|---|---|---|
| May 28, 1975 | 77 | 76 |

| 8a. CONTRACT OR GRANT NO. | 9a. ORIGINATOR'S REPORT NUMBER(S) |
|---|---|
| N00017-C-73-1418 | |
| b. PROJECT NO. | TM-75-165 |
| c. N00017-73-C-1418 | 9b. OTHER REPORT NO(S) *(Any other numbers that may be assigned this report)* |
| d. | |

10. DISTRIBUTION STATEMENT

~~Each transmittal of this document outside the Department of Defense must have prior approval of the Naval Air Systems Command, AIR 370F.~~

| 11. SUPPLEMENTARY NOTES | 12. SPONSORING MILITARY ACTIVITY |
|---|---|
| | NAVAL AIR SYSTEMS COMMAND |

13. ABSTRACT

This report reviews relevant knowledge in the context of passive sonar aural recognition of a noise source, and defines an experimental approach to extend the state of knowledge in areas of immediate concern to the Naval Air Systems Command. The report is prepared in two parts with Part 1 summarizing the sonar classification task and then leading into the relationships of this task to the general topic of acoustic warfare.

Part 2 of the report considers the aural classification task in more general terms. The results of studies using trained listeners are presented and are compared to those predicted from previous studies done elsewhere. A model of the classification task as a specialized detection problem is presented and it is shown that the model allows prediction of the results within a few dB of signal-to-noise ratio required to make a terminal classification decision.

| 14. KEY WORDS | LINK A | | LINK B | | LINK C | |
|---|---|---|---|---|---|---|
| | ROLE | WT | ROLE | WT | ROLE | WT |
| ACOUSTIC WARFARE | | | | | | |
| AUDITORY RECOGNITION | | | | | | |
| AURAL CLASSIFICATION | | | | | | |
| PASSIVE SONAR CLASSIFICATION | | | | | | |
| PATTERN DISCRIMINATION | | | | | | |
| PSYCHOACOUSTICS | | | | | | |
| SEQUENTIAL RECOGNITION | | | | | | |
| SIGNAL DETECTION | | | | | | |

ACCESSION for

NTIS          White Section ☐
DDC           Buff Section  ☐
UNANNOUNCED                 ☐
JUSTIFICATION  Per
letter on file

BY
DISTRIBUTION/AVAILABILITY CODES

Dist          SPECIAL

A

ABSTRACT

This report reviews relevant knowledge in the context of passive sonar
aural recognition of a noise source, and defines an experimental approach to
extend the state of knowledge in areas of immediate concern to the Naval Air
Systems Command. The report is prepared in two parts with Part 1 summarizing
the sonar classification task and then leading into the relationships of this
task to the general topic of acoustic warfare.

Part 2 of the report considers the aural classification task in more
general terms. The results of studies using trained listeners are presented
and are compared to those predicted from previous studies done elsewhere. A
model of the classification task as a specialized detection problem is pre-
sented and it is shown that the model allows prediction of the results within
a few dB of signal to noise ratio required to make a terminal classification
decision.

TABLE OF CONTENTS

## LIST OF TABLES

## LIST OF FIGURES

## CHAPTER 1

## INTRODUCTION

Many of the noises to which man is exposed serve also to convey information about the environment in which he lives. The noises serve as a masking stimulus which decreases his ability to use or enjoy desirable signals such as speech signals or music. In addition, however, the ever-present noise stimuli can convey useful information and under proper conditions such noises allow the listener to identify the source and to make inferences about it.

The role of noise waveforms is of considerable significance in a number of situations ranging from the need of a submarine to detect and identify hostile or otherwise potentially hazardous 'targets' to that of a lathe operator's ability to adjust the cutting rate of his machine. The ability to identify sources of ocean-borne noise is of great concern to the Navy and, in spite of ever more sophisticated special purpose processing machines, the role of the sonar operator in this context is still significant. Recently Urick and Gaunaurd (1972), Stallard and Leslie (1974) have addressed the question of psychophysics in sonar detection, and a considerable body of experimental work has dealt with the related question of parameters important in speech recognition (Stevens and House, 1972.)

The objective of this report is to review relevant knowledge in the context of passive sonar aural recognition of a noise source, and to define an experimental approach to extend the state of knowledge in areas of immediate concern to the sponsor, the Naval Air Systems Command (NAVAIR.) The exposition which follows will touch on a number of areas to include auditory detection, learning, response uncertainty, and subject selection and instruction. The results of pilot studies using trained college age listeners are also presented.

The report is prepared in two parts with part 1 summarizing the sonar classification task and then leading into the relationships of this task to the general topic of acoustic warfare. Part 2 considers the classification

task in more general terms.  The results obtained by a number of experimenters
in psychology are discussed in their relationship to the present effort.
Also, various statistical techniques and controls exercised during
experiments done to date are presented.

## CHAPTER II

## THE AUDITORY RECOGNITION TASK

### 2.1 General

In this chapter the problem posed by the need of the listener to identify or recognize an auditory source is investigated. Such recognition first implies that the listener has somehow been convinced of the presence of the source and must now decide to which of a number of possible classes of sounds the source belongs. Indeed in practice the processes of detection and classification are closely related as will be seen in what follows. Also, in most cases of practical interest the source will be heard in the presence of some extraneous noises which will tend to mask the presence of the source as well as confound the recognition or classification task.

To cast the problem in more concrete terms, consider the case of a sonar operator using aural means only and facing the prospect of receiving the acoustic emanations of only a single source, a target. We may define the task precisely using a parallel to the general estimation model (Van Trees, 1968) modified to include some concepts from pattern recognition (Fu, 1968.)

[a] A source, the target, may for our purposes be classified by a set of acoustic parameters. These parameters $\bar{a}$ define a multi-dimensional parameter space $\nabla$ and they may be associated with such features as the absence or presence of certain bands of noise, the modulation characteristics, or the time dependence of the sound. We will denote the pattern corresponding to source j as $a_j$ and the features as $\Omega_j$, see Figure 1.1

[b] The source parameters are mapped into an observation space $\theta$ according to some probabilistic law. The observable $\bar{R}$ is in our case the listeners response to the stimulus. The observation, as perceived by the listener, includes the effects of the auditory transducer as well as the subsequent neural and higher level cognitive processes.
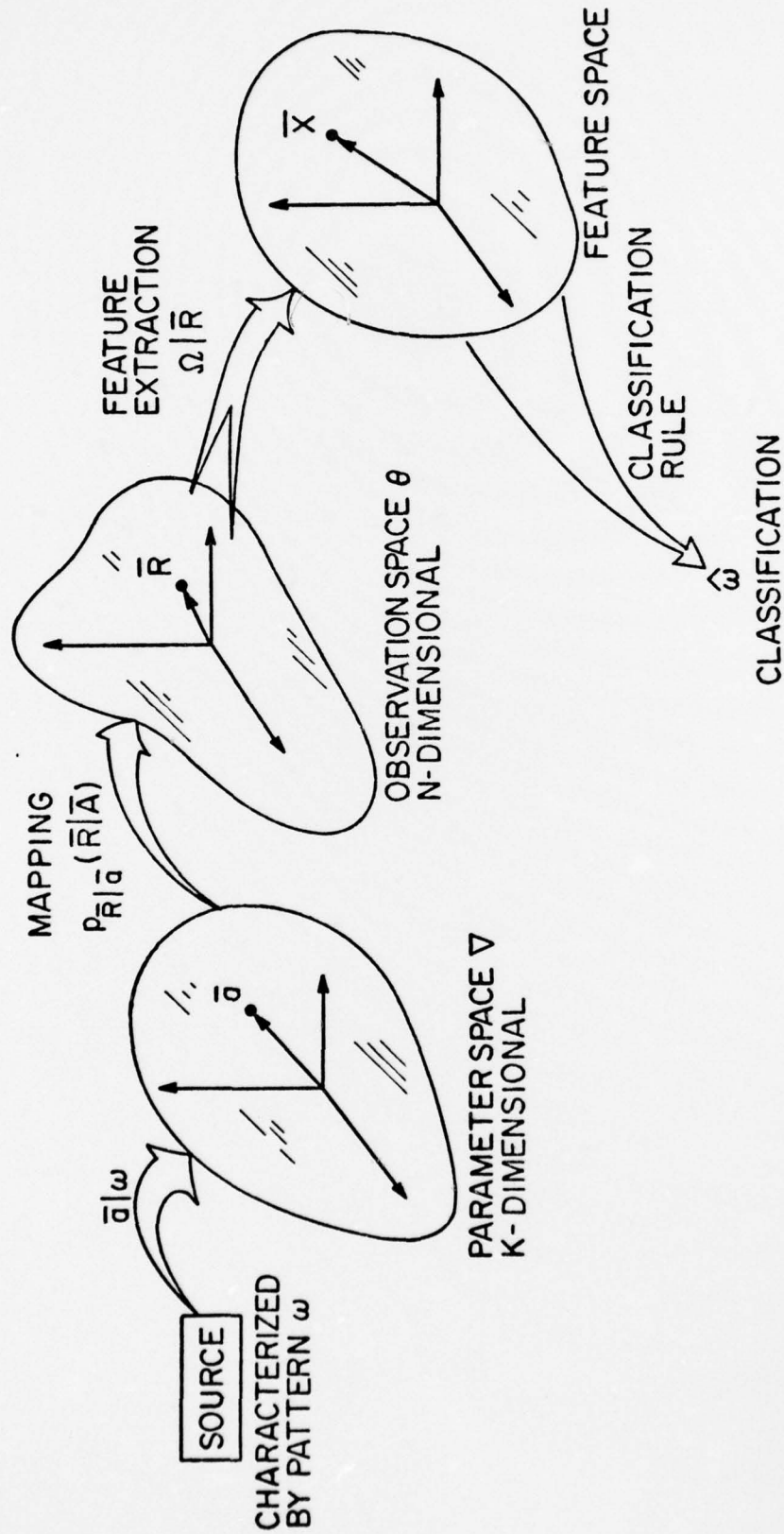
Figure 1.1  General Classification Model

[c]  It is suggested that the listener perceives the set of
     observables as features and organizes these into patterns.
     An  N  dimensional observation space can then be reduced
     to an  I  dimensional feature space.  As an example of this
     process, we can consider the critical band concept which
     successfully explains many auditory phenomena.  (Plomp and
     Smoorenburg 1970, Zeicker, Flottorp, and Stevens, 1957.)

[d]  It is further proposed that the listener will try to compare
     the pattern he hears to one of a number of patterns previously
     learned in order to arrive at an identification of the source
     $\hat{\omega}$ $(\bar{R})$.

In any realistic listener classification task, the patterns $\omega_j$, the
features, and hence parameters $\bar{a}$ occur with some a priori probability.  The
likelihood of a specific pattern strongly influences the classification
problem by restricting the number of patterns which the listener expects to
hear.

In the above model of the recognition task, the listener is seen to
perform the functions of feature extraction and pattern classification.  The
probabilistic mapping from the parameter to the observation space includes
such perturbations as are introduced by the presence of noise either between
the source and the listener's auditory end-organ or subsequently in auditory
pathways.

## 2.2  Patterns with Dichotomous Features

As a foundation from which to proceed to consider realistic recognition
tasks and which may allow for some psychoacoustic tests of the concepts, we
now consider the case of dichotomous features.  Specifically, we wish to
treat the case where source acoustic patterns differ by either having or not
having specific features.

To further simplify, consider the recognition task where the listener
must decide between two signal patterns $\omega_1$, $\omega_2$ only, and they differ by the
presence or absence of a single feature $\Omega_i$.  The listener then is faced
with the following choices:

a)  If the perceived pattern is thought to have feature $\Omega_i$, say that
    it is pattern $\omega_1$.

b) If feature $\Omega_i$ is thought to be missing, say that the pattern heard is pattern $\omega_2$.

In this case the listener is then faced with the task of deciding between two hypotheses:

$H_0$: Feature $\Omega_i$ is absent, decide pattern $\omega_2$

$H_1$: Feature $\Omega_i$ is present, decide pattern $\omega_1$

on the basis of a sensural input.

If feature $\Omega_i$ were the only feature of the signal, the problem would reduce to that of detection of a signal. While the detection problem is certainly of interest, in the present work we are rather concerned with distinguishing between pattern classes. The detection problem has been treated elsewhere (Andrews and Hovater, 1971, Urick and Gaunaurd, 1972) but it is instructive in the present case to state some of the applicable concepts.

It is reasonable to suspect that the decision between $H_0$ and $H_1$ requires at the very least an opportunity for the listener to <u>detect</u> feature $\Omega_i$ when hypothesis $H_1$ applies. If in fact this detection opportunity is sufficient for making the decision is as yet an open question. In attempting to analyze the problem in more detail, the theory of signal detectability will be reviewed. Also, it will be seen later that in some experimental cases the detection of dichotomous features seems to be an adequate basis for a classification decision.

## 2.3 The Theory of Signal Detectability

The theory of signal detectability is a mathematical model of the mapping of a set of stimuli, consisting either of noise alone or from signal plus noise, into a two-point decision space: "no" the signal was not contained in the input, or "yes" the signal was contained in the input (Tanner and Sorkin, 1972.) While these models are not intended to be descriptions of the way a human observer makes this binary decision, they are normative models with which the observer's performance can be compared.

The models comprising this theory are applicable to any receiver operating on a set of inputs and as such, do not take into account the properties of the human per se.  A number of workers have successfully demonstrated the utility of the theory of signal detectability of visual as well as auditory stimuli (Green and Swets, 1966, Peterson, Birdsall and Fox, 1954, Swets, 1964.)  It is generally found that the human performance falls short of that predicted for an ideal detector.

The radiated noise of a marine target consists of a broad band noise spectrum with perhaps some significant tonals associated with machinery on board.  The problem of detecting the presence of various broadband components reduces to a test of hypotheses about signals which are only known statistically. That is, the exact phases of the components, their instantaneous amplitudes, bandwidths, etc. are at best known in terms of their averaged power spectra. The theory of signal detectability for broadband features leads to a likelihood ratio test for signals known statistically.  If we further assume that the radiated noise can be modeled as a collection of samples with identical Gaussian probability densities, the ideal processor is an energy detector (Van Trees, 1968.)

From the sampling theorem in the time domain, a band-limited signal with bandwidth W can be exactly determined by 2WT samples.  These samples will be statistically independent if the process is Gaussian (Lathi, 1968.)  The maximum log likelihood ratio test is then given by

$$\chi(\bar{R}) = \sum_{i=1}^{2WT} R_i^2 \underset{H_0}{\overset{H_1}{\gtrless}} \gamma = L(R^2)$$

where the threshold $\gamma$ includes all sample invariant factors as well as the criteria.  (Van Trees, 1968.)  We can see that the likelihood ratio test consists of computing the sum of squares of the statistically independent sample data points and comparing this to an appropriate threshold.  Since the $R_i$'s are Gaussian random variables, the sufficient statistic $\chi(\bar{R})$ is a random variable with a gamma distribution

$$P_\ell(\chi) = \frac{1}{\Gamma(\alpha)\beta^d} \, \chi^{d-1} e^{-\chi/\beta} , \; 0 < \chi < \infty. \qquad (2.1)$$

In this case

$$d = \nu/2 = WT, \quad \beta = 2\sigma_h^2 \quad \text{(Hogg, 1970)}.$$

The parameter $\sigma_h^2$ is given by:

$$\sigma_h^2 = \sigma_n^2 \text{ under } H_o \text{ if } R \text{ is } n(o, \sigma_n^2)$$

$$\sigma_h^2 = \sigma_n^2 + \sigma_s^2 \text{ under } H_1 \text{ where } R \text{ is } n(o, \sigma_n^2 + \sigma_s^2)$$

The notation $n(\mu, \sigma_s^2)$ is used to denote a random variable with Gaussian probability density which has a mean $\mu$ and a variance $\sigma^2$. Equation (2.1) reduces to that of a chi-square density function with $\nu$ degrees of freedom whenever $\sigma_h^2 = 1$.

Furthermore, whenever the number of degrees of freedom

$$\nu = 2WT > 100 ,$$

the random variable $\chi(R^2)$ can be closely approximated by a Gaussian density with the following parameters under $H_o$ and $H_1$ (Abramowitz and Stegun, 1964)

$$P_{\chi|H_o} (X|H_o) \text{ is } n(2WT \, \sigma_N^2, \, 4WT \, \sigma_N^4)$$

$$P_{\chi|H_1} (X|H_1) \text{ is } n(2WT[\sigma_n^2 + \sigma_s^2], \, 4WT[\sigma_n^2 + \sigma_s^2]^2) .$$

In keeping with common usage (Swets, 1964), define a detectability index as:

$$d'_{opt} = \sqrt{2} \left\{ \frac{E(X|H_1) - E(X|H_o)}{1/2[\text{Var } (X|H_1) + \text{Var } (X|H_o)]} \right\}^{1/2}$$

$$= (WT)^{1/2} \frac{\sigma_s^2}{\sigma_n^2[1/2(\sigma_s^2/\sigma_n^2)^2 + (\sigma_s^2/\sigma_n^2) + 1]^{1/2}} . \quad (2.2)$$

For the small signal-to-noise case where

$$\sigma_s^2/\sigma_n^2 \ll 1 ,$$

Equation 2.2 reduces to the usual case where the variance under the two hypotheses is the same. Then

$$d'_{opt} = (WT)^{1/2} \; \sigma_s^2/\sigma_n^2 \;, \quad \sigma_s^2/\sigma_n^2 \ll 1.$$

for a forced choice test.

## 2.4 Deciding Between Complex Patterns

In the preceding sections we assumed that a minimum requirement for distinguishing between two classes of patterns was the opportunity to detect the distinguishing features. We then went on to define a measure of detectability, the detectability index $d'_{opt}$, for bands of noise. An equivalent measure of detectability can in theory be written for all features which characterize the patterns. Actual marine noise sources are much more complex than this, however, and one wonders about the utility of these simple models. In an attempt to relate the problems, below is an ordering of auditory recognition tasks from simple to operationally realistic.

a) Distinguishing between two patterns differing by one dichotomous feature.

b) Distinguishing between two patterns differing by a number of dichotomous features.

c) Distinguishing between two pattern classes which differ by one feature which can exhibit a continuous range of detectability or feature parameters.

d) Distinguishing between two pattern classes which differ by a number of features with ranges of detectabilities.

e) Deciding to which of N pattern classes a sound belongs when the sound has a number of features with ranges of detectabilities or parameters for each.

We can see that the problem with which the sonar operator is faced is either d or e above. Whenever the operator is somehow convinced that the a priori probability of targets is such that only two classes of patterns are to be expected, the problem is simplified but still is a very complex one.

Also, nothing has been said so far about his degree of knowledge (learning) about the patterns with which he is attempting to match the unknown.

Whenever a number of features are associated with a pattern, it can be expected that some features will occur only when others are present. The features are therefore correlated one to another to some extent. As an example of this idea, consider the pattern classes consisting of:

$\omega_1$: a low speed surface ship

vs $\omega_2$: a surface ship proceeding at a speed high enough to cause prop cavitation.

The pattern class $\omega_1$ will exhibit a number of tonals which are associated with machinery on board. Also, the sound will be only lightly amplitude modulated by the effect of waves on the acoustic coupling to the water. As the ship increases speed, some tonals will shift in frequency while others remain stable. Also, the broadband noise will increase and will tend to mask the tonals. Once prop cavitation occurs, there is a speed range over which the broadband sound will be strongly amplitude modulated. Other changes also occur in the sound depending on the type of ship, the number of screws, etc. (Myasnikov and Myasnikova, 1971.) How do we attack this complex problem in an efficient manner?

One approach is to present members of the two pattern classes to listeners under various listening conditions and to elicit classification responses. From such experimental results we can obtain a measure of operator performance for a specific set of pattern classes under the conditions tested. This method has, in fact, been applied to a number of sonar cases as well as to speech perception (National Research Council, 1949, Licklider, 1951, Pollack, 1948.) Based on a large number of such experiments, we can perhaps extend the observed results to classification of as yet untested sources.

Alternately, we can treat simple cases with an attempt to model the process. Hopefully we can then infer the performance in the more complex cases. Certainly some experimental verification with these complex sounds would be needed if the model is to be useful. Listed below are some of the techniques which can be considered in the extension of simple concepts to operationally significant ones.

a)  Inter-relationships between features which are correlated can
    be reduced to second-moment characterizations in terms of
    uncorrelated random variables using the Karhunen-Loève expansion
    (Van Trees, 1968, Fu, 1968.)

b)  These orthogonal feature eigenvectors can then be treated using
    detection theory in the case of dichotomous features or
    estimation theory when features can occur over a range of
    detectabilities or parameter values.

c)  Classification into one of N pattern classes can then be modeled
    using the concept of discriminant functions and other pattern
    recognition procedures (Fu, 1968, Sholl, 1971.)

## CHAPTER III

## SEQUENTIAL RECOGNITION

### 3.1 Role of Criteria and A Priori Knowledge in Sequential Detection

In keeping with the assumptions made in the foregoing chapter, we will now go on to analyze in greater detail the detection of a dichotomous feature, $\Omega_i$. Operationally the sonar operator is faced first with the problem of detecting a target and then with the classification of that target. This is a sequential detection problem in both the initial detection case and the detection of the feature $\Omega_i$.

At any time $t_j$, the listener is faced with three possible choices:

a) On the basis of the aural stimulus, decide $H_o$, the feature $\Omega_i$ is absent.

b) Given the aural stimulus, decide $H_1$, the feature $\Omega_i$ is present.

c) Decide to continue listening because there is not sufficient evidence to decide either $H_0$ or $H_1$.

The listener will choose his course of action depending both on the information in the input stimulus and on his mental set (Leeper, 1951.)

To gain insight into the effect of a priori knowledge and response criterion, consider the two cases below. In both cases the two pattern classes to be decided between are $\omega_1$, the pattern associated with a light warship and $\omega_2$, the acoustic pattern of a tanker.

a) A modern tanker can have a draft exceeding 100 feet and the submarine is required to transit at rather shallow depth in proximity to a heavily traveled shipping lane.

b) In times of hightened military tension, the submarine is to maintain covert surveillance of all military surface ships in a strategic sector.

In the first case, the penalty for failing to classify the tanker could be very great, whereas misclassifying the warship as a tanker would only lead

to a minor penalty due to having to skirt a potentially harmless shallow draft craft. Also, the operator would be anticipating merchant shipping. In the second case, the situation is less clear cut since misclassifying a tanker as a warship would dilute the efforts of the submarine. Falsely dismissing the warship would also be a serious error, however.

Whenever the a priori probabilities of the signals are known, and a cost can be associated with the available courses of action, the Bayes criterion can be used to minimize the total risk (Van Trees, 1968.) In the sequential detection problem we then have:

$P_1$, $P_2$, the a priori probabilities of $\omega_1$, $\omega_2$ respectively.

$C_{ij}$, i, j = 1,2 where $C_{ij}$ is the cost of classifying the pattern as $\omega_i$ given that $\omega_j$ is true.

$C_{def}$, the cost of deferring the decision until more information is available.

Alternately, if the a priori probabilities are not known, or if it is not reasonable to assign costs to each of the courses of action, it may be more appropriate to set a criterion which reduces the probability of, say a false classification as $\omega_1$, below some limit. The threshold adopted in this case will be chosen using the Neyman-Pearson criterion.

Whatever the criterion by which the thresholds for the available courses of action are determined, the observed performance will be strongly influenced by this threshold. A way to characterize this performance under various criteria is by means of a receiver operating characteristic (ROC) curve. This curve presents the probability of a detection P(D) vs the probability of a false alarm P(FA). As the listener adjusts his criteria, both of these probabilities will change.

Under a lax criterion, i.e., say $\omega_1$ whenever there is any evidence of the presence of feature $\Omega_i$, the number of correct detections will be high as will the probability of a false alarm. A more strict criterion will decrease both of these probabilities.

## 3.2 Factors Influencing the Terminal Decision

Whenever a decision to accept hypothesis $H_o$ or $H_1$ is made, we refer to that as a terminal decision. How is such a terminal decision made? The initial target detection has been treated elsewhere, and will not be reiterated (Loeb, 1970, Boehme, 1970.) But, knowing that a source is present, the listener must now decide about the presence of a specific feature.

If we assume for a moment that in each detection opportunity the decisions of $H_o$ and $H_1$ both have finite probabilities, it is only a question of time until a terminal decision will occur by chance alone.

Wald (1947) extended the concept of the likelihood ratio test to the sequential detection problem. The sequential probability ratio test is, at the i'th stage

$$B \overset{H_1}{<} \lambda_i(\bar{R}) = \prod_{j=1}^{i} \frac{P_{r|H_1}(R_j|H_1)}{P_{r|H_o}(R_j|H_o)} \overset{H_o}{<} A \quad, A < B$$

else defer the decision.

Here A and B are the thresholds for deciding $H_o$ and $H_1$ respectively. He goes on to prove that such a test will always terminate and he derives the expected number of observations for a terminal decision.

In the passive sonar case we must be very specific, however, in the way we define a detection opportunity or observation. Furthermore, the above development assumes that the probability densities under $H_o$ and $H_1$ remain unchanged from observation to observation. The actual detection occurs in a situation where the source to receiver range, hence signal-to-noise ratio and probability densities, change as a function of time. The criteria and associated threshold are also likely to change in the real-world situation. If, for example, the time required to come to a decision is very long, the criterion may be relaxed for the sake of arriving at a decision so that another target can be investigated.

## 3.3 Forced Response vs. Sequential Classification

Classifications which imply only the detection of one dichotomous feature may be conveniently visualized by the probability densities under $H_o$ and $H_1$.

Whenever the listener must either make a yes-no (YN) or a forced choice, 2 alternative (2AFC) response a single threshold divides the decision space into two distinct regions. Figure 2.1a shows the probability densities and error regions for this situation for the detection of a band of noise in noise. The error regions are:

$$\alpha = P(\text{deciding } H_1 | H_o) = P(FA) = \int_\gamma^\infty p_{\chi | H_o} (\chi | H_o) \, dx$$

$$\beta = P(\text{deciding } H_o | H_1) = 1 - P(D) = \int_o^\gamma p_{\chi | H_1} (\chi | H_1) \, dx.$$

The error regions at time $t_i$ for sequential detection are shown in Figure 2.1b. The probability of an error in classification will be given by:

$$P(E)_t = P(H_o) \, P(\chi | H_o)_t + P(H_1) \, P(\chi | H_1)_t$$

$$= \alpha_t \, P(H_o) + \beta_t \, P(H_1)$$

where

$$\alpha_t = \int_{\gamma_B}^\infty p_{\chi | H_o} (X | H_o) \, dx, \quad \beta_t = \int_o^{\gamma_A} p_{\chi | H_1} (X | H_1) \, dx.$$

In practice, the criterion, hence threshold, will be set by the listener in response to the situation. The numeric value of the threshold $\gamma$ will not be known but in the forced response case can be inferred from the observed values of $P(FA)$ and $P(D)$. The sequential detection and classification case presents some difficulties, however. Especially when the distributions under $H_o$ and $H_1$ are changing, knowledge of $\alpha_t$ and $\beta_t$ is itself not adequate to define $\gamma_A$ and $\gamma_B$ without assumptions of the terminal decision process.

The points made in this chapter may appear of only theoretic interest. However, we will see that these are central issues in the design of psycho-acoustic experiments of aural classification.
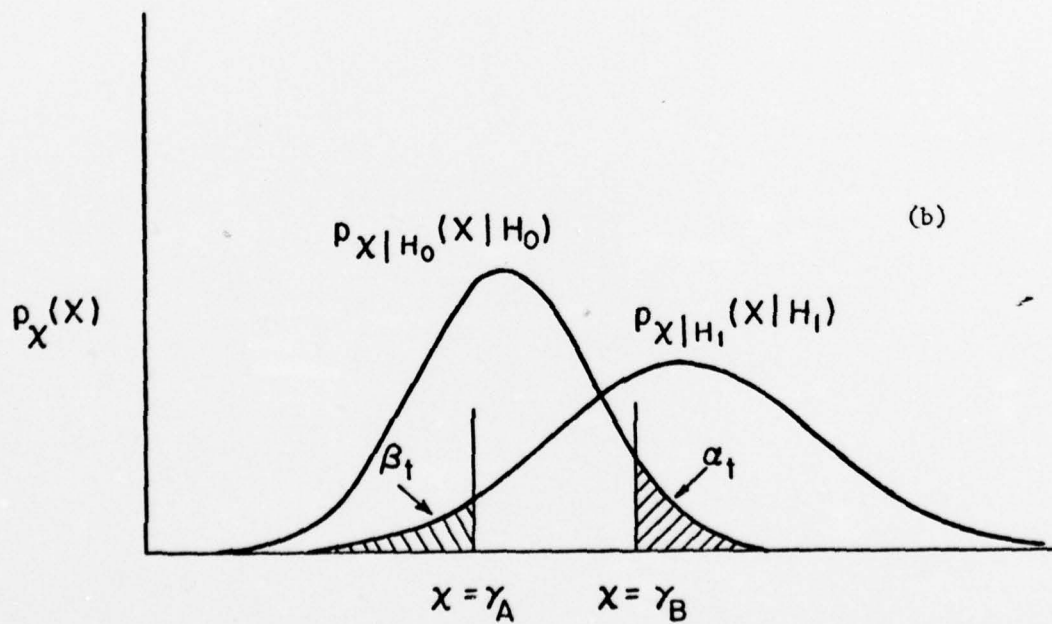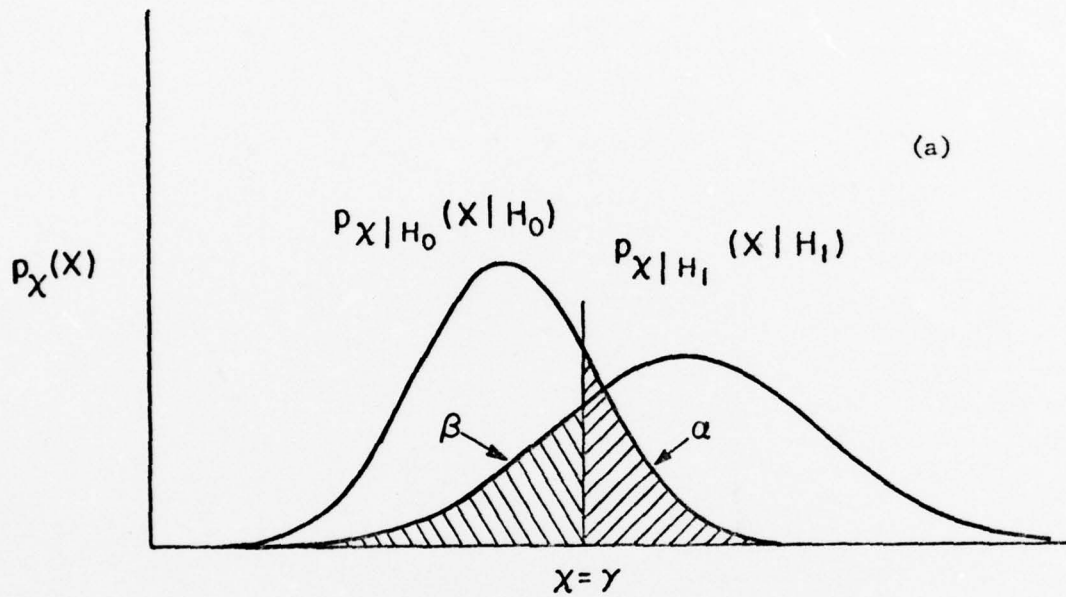
Figure 2.1  Error Regions for Forced Response and
Sequential Detection

## CHAPTER IV

## DESIGN OF AN EXPERIMENT IN AURAL RECOGNITION

### 4.1  Threshold or Forced Response Tests?

Classically, the basic correlates of the auditory stimulus have been investigated by threshold methods (Boring, 1950.)  In this approach, the stimulus parameter to be measured is changed slowly until a just perceptible difference in the physiological response occurs (Licklider, 1951.)  Such techniques, with variations in details of implementation, have been used to investigate the equal loudness of various spectrally shaped noises, for example.  The response was one of "louder than" or "not as loud as" a reference band of noise (Cremer, Plenge, and Schwarzl, 1959, Zwicker, 1958, Zwislocki, 1969.)

The results of these threshold experiments can be explained using the theory of signal detectability.  It can be inferred that the listener operates in a sequential detection or recognition environment.  Whenever the established criterion is exceeded during a period of increasing stimulus intensity, the threshold can be thought to have been exceeded.  Also, many of these experiments alternately increase and decrease the stimulus intensity with the aim of "bracketing" the threshold value.

It should be pointed out, however, that single or multiple thresholds for sensural inputs have been postulated as explaining the observed results.  Also, quantization of the auditory process has been suggested (Licklider, 1951.)  While we are not concerned here with the analysis of detailed mechanisms which are active in the human organism, there are basic differences in approach between experimenters using threshold techniques and those testing various aspects of the signal detection theory.  These differences in technique are of considerable importance.

Detection theory related experiments are designed to elicit either data about the ROC curve or the psychometric function of the listener.  The psychometric function is the variation of percent correct detection or classification as a function of stimulus level.  The key parameters are usually the various probabilities of the available courses of action and the detectability $d'_{opt}$ expected on the basis of some signal detection model.  The detectability is usually defined as the normalized difference of means of a statistic under two alternate hypotheses.

The signal detection theory types of experiments present the stimuli to a listener at some signal-to-noise ratio and observe the performance. The criterion is usually specifically included by way of instructions prior to the test.  Data are taken at a number of signal-to-noise ratios or other fixed values of a parameter (Swets, 1964.)  The responses are either "yes" the signal was present in a trial period, or "no" it was not, a YN type experiment, or a forced choice N alternative NAFC experiment where the listener indicates in which one of N trial periods the signal occurred. Variations of these methods are also used, and the signals are usually presented in the presence of a noise at some preselected level.

The latter experimental technique, the YN or NAFC tests, has the advantage that the results are subject to well known and easily applied statistical hypothesis testing techniques as are described in Chapter II.

## 4.2  A Pilot Study and its Impact

Initially, a pilot study was set up using university students as subjects to identify weaknesses in experimental techniques and to obtain bounds on listener performance.  This was an identification task in which the listener was asked to decide which of two trained marine sounds (the exposure set) was subsequently presented in a broadband noise background. The signals were presented binaurally from tape recordings in an experimental arrangement very similar to that which is described in later sections. However, for these initial tests the unknown or probe stimulus was presented for a 20 second period of time at a fixed signal to noise ratio.  The listener was asked to record on an answer sheet his or her assessment of which member of the exposure set the probe corresponded to.  In addition, the third alternative "don't know" was to be indicated whenever "reasonable" doubt existed.

Because the sounds presented do not fall within the range of normal auditory experience for these listeners, an immediate area of difficulty arises. We are precluded from using techniques applicable to, for example, speech intelligibility studies where the subjects can be assumed to have a common learned ability to distinguish between signals (Pollack, 1948). A time consuming stimulus training period is required prior to any probe measurement.

Also, since the stimuli characteristics cannot be reliably verbalized, an arbitrary designation of A and B were associated with the trained stimuli. Such arbitrary association of letters with the members of the exposure set is unfortunate. However, the methods used by Pollack (1959) to circumvent this shortcoming, the method of recognition memory, cannot be used in our case. In that method, the probe is chosen from an augmented set which includes the members of the exposure set and members of another set, the confusion set. The listener response then consists of an assessment of whether the probe in fact corresponded to a member of the exposure set. The marine sounds are such that the experimenter cannot know if the confusion signals will not, in fact, sound "like" one of the exposure signals. The listener cannot establish a mental "measure" of the dissimilarity of signals in this case.

A test event consists of an exposure (learning) period for the two signals without interfering noise. After a short pause, the probe is presented in the interfering noise. Subsequently, the exposure set is repeated in a refresh period, or a new set of signals is exposed for learning. No feedback was used except that in some cases the subjects were appraised of their performance at the end of a test session. Each test event requires about two minutes to complete for a total of 15 to 18 events per session.

Little was known at the time of this pilot study about the range of signal-to-noise ratios at which the listener performance would attain some pre-determined probability of correct response $P(C)$. It was necessary, therefore, to present the probe at a number of values of signal-to-noise ratios. Using four signal pairs of interest and five values of signal-to-noise ratio, there were 20 performance indices to be estimated. As will be shown later, one would prefer at least 50 data points per performance index for a total of 1000 events or about 60 sessions.

The availability of trained listeners willing to participate in these studies was limited, and not all performance indices were adequately studied. This method was abandoned in favor of the ramped signal-to-noise ratio technique discussed in subsequent sections of this report primarily because a faster method was required. This early study was very useful in identifying several weaknesses in procedures and equipment.

## 4.3 A Modified Threshold Procedure

Most of the results presented in this report were collected using a modification of threshold procedures. The scheme used is modeled after the techniques used by Cremer (1959) and Zwicker (1957) to determine the loudness of various bands of noise. Some considerations which strongly influenced the experimental techniques are listed below.

a) Because it was desired at the outset to make the listener-related portions of the system portable so as to be useable at the sonar schools, the signals are pre-recorded on audio tape. Also, the response recorder is of necessity simple.

b) Responses needed to be recorded automatically without the test conductor needing to be present because of other demands on the author which required considerable periods of absence from the University.

c) The process of data reduction was to be as simple as practical so that no one individual would be overly committed to this portion of the effort.

In the modified threshold procedure, the probe is initially presented in a broad band noise background at a low signal-to-noise ratio. This signal-to-noise ratio (SNR) is then increased linearly with time until such time as the listener can make a determination of which exposure signal the probe corresponds to. This method has an advantage in its similarity to the sequential classification task faced by a sonar operator in a real-world decreasing range situation. The results only provide data about the probability of a correct (or wrong) classification whenever the listener has sufficient confidence to make a terminal decision, however.

The sequence of occurrences during each test event is shown in Figure 4.1. The exposure signals A and B are chosen with randomization from the set of auditory patterns of interest. Generally, these patterns differ by one or more features. The initial and refresh exposures are of fixed duration which the listeners generally agreed on as sufficient for later recognition. The probe SNR is always low initially but this value is randomized to avoid listeners responding on the basis of time instead of the perceived aural response.
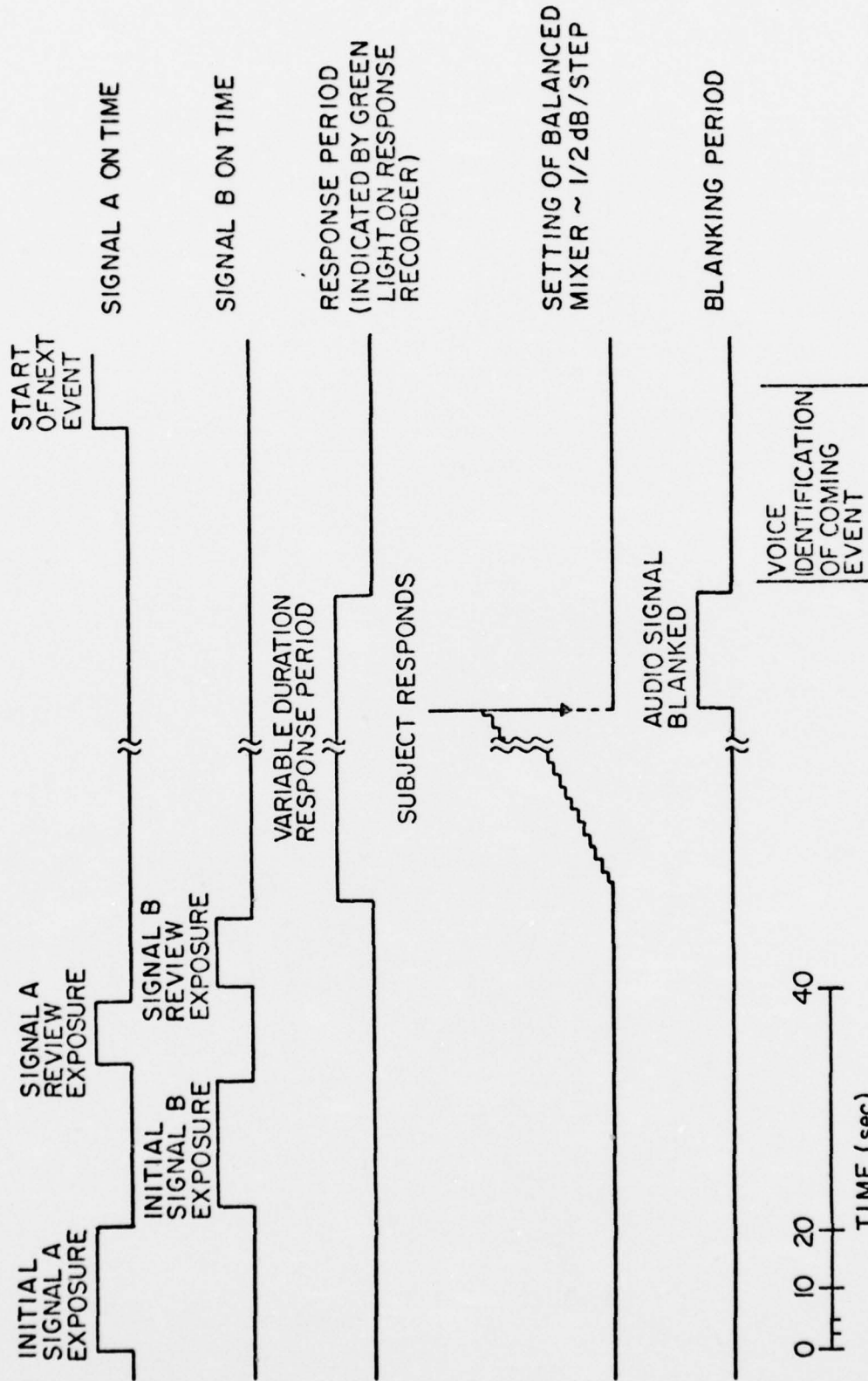
Figure 4.1 Overall Timing Diagram for an Event of the Modified Threshold Procedure

Responses were recorded by listeners by pressing either a switch marked
A or one marked B. Once the listener indicated a decision, the auditory signal
was blanked for the remainder of the response period. The duration of the
response period was also randomized. By not knowing when the response period
would end, the listeners were in effect assigned a cost for continuing in the
listening state. Initially, it was planned to compensate the listeners on
the basis of performance. The payoff would have been computed on the basis of
correct classifications with a penalty for wrong answers, and no payoff for
events where no classification was made. The use of trained, expert listeners
obviated the need for this gaming approach because the level of motivation
was high and the level of performance could be established by verbal instructions
alone.

The pre-recorded signals were played back using a Crown 700 tape player
and a pair of Telephonics Model TDH-39 dynamic earphones. The signal was
fed in phase to a matched set of these headphones and the listener was further
acoustically isolated by being enclosed in an audiometric booth. Figure 4.2
diagrams the equipment needed at the listener location. A Sony cassette
recorder is used for recording the responses which are coded as frequencies.
Also recorded onto this cassette is a tone which is proportional to the SNR.
This tone and other signals needed to properly sequence the response equipment
are recorded on one channel of the two-channel Crown tapes. The exposure, probe
and interfering noise signals are recorded on the other channel of these tapes.
The equipment needed at the site of the tests is pictured in Figure 4.3.

To analyze the data, the cassette tapes were played back on another cassette
tape unit. The SNR related tone and the response tones were then output on a
cash register tape by means of a counter and printer combination. Manual methods
were used to determine the SNR at the moment when the listener responded. Also,
the response (A or B) and the time from the beginning of the response period
were recorded on a data form. As the number of data points accumulated, the
results were key punched and copied onto a computer magnetic tape for retention
and accessability. Computer software to perform various statistical analyses
on the data has also been developed.

4.4  Listener Safeguards and Calibration Techniques

Prior to beginning experiments using human subjects, approval was obtained
in compliance with the Institutional Assurance provisions of The Pennsylvania
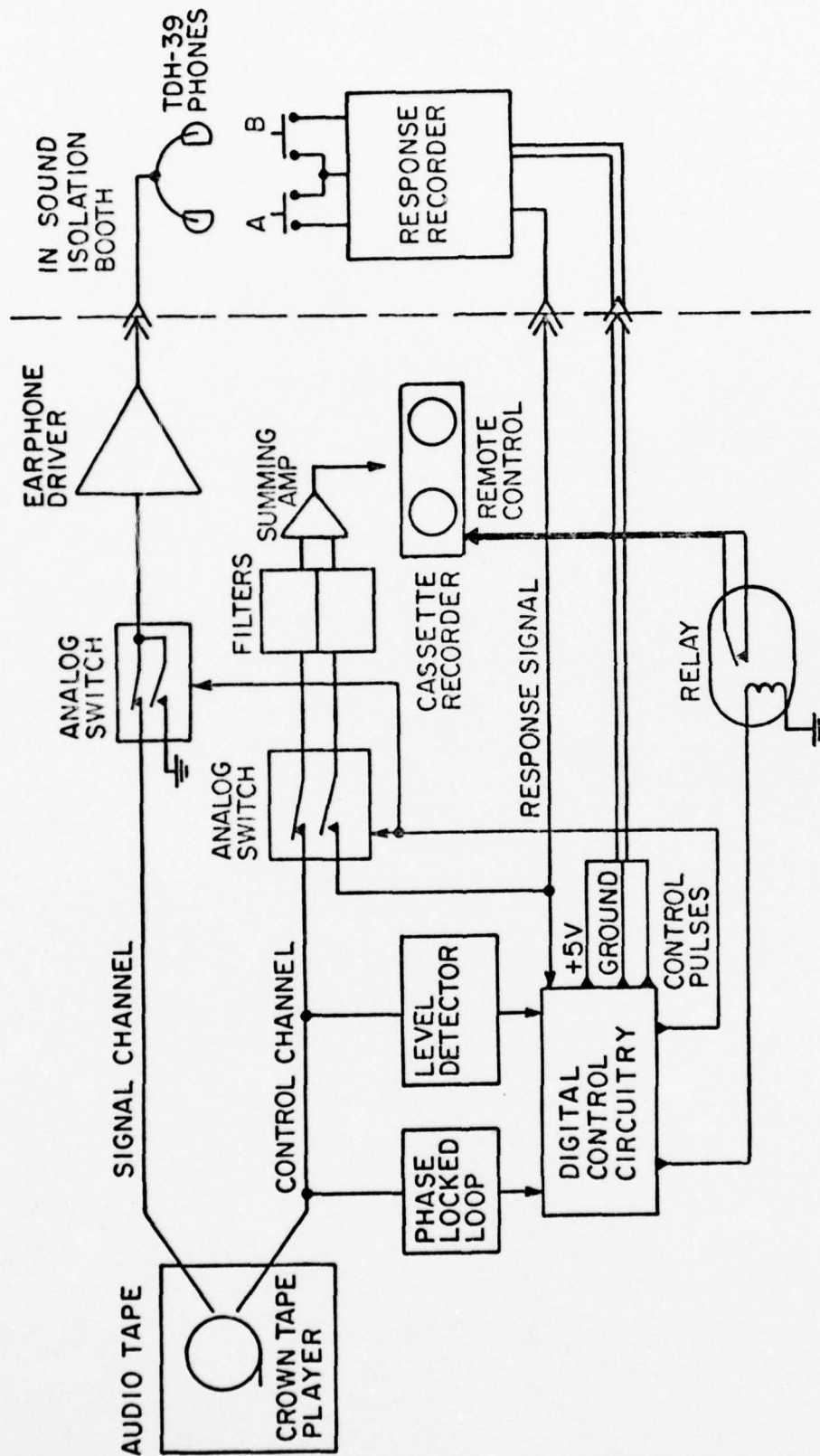
Figure 4.2 Summary of Equipment of the Test Site
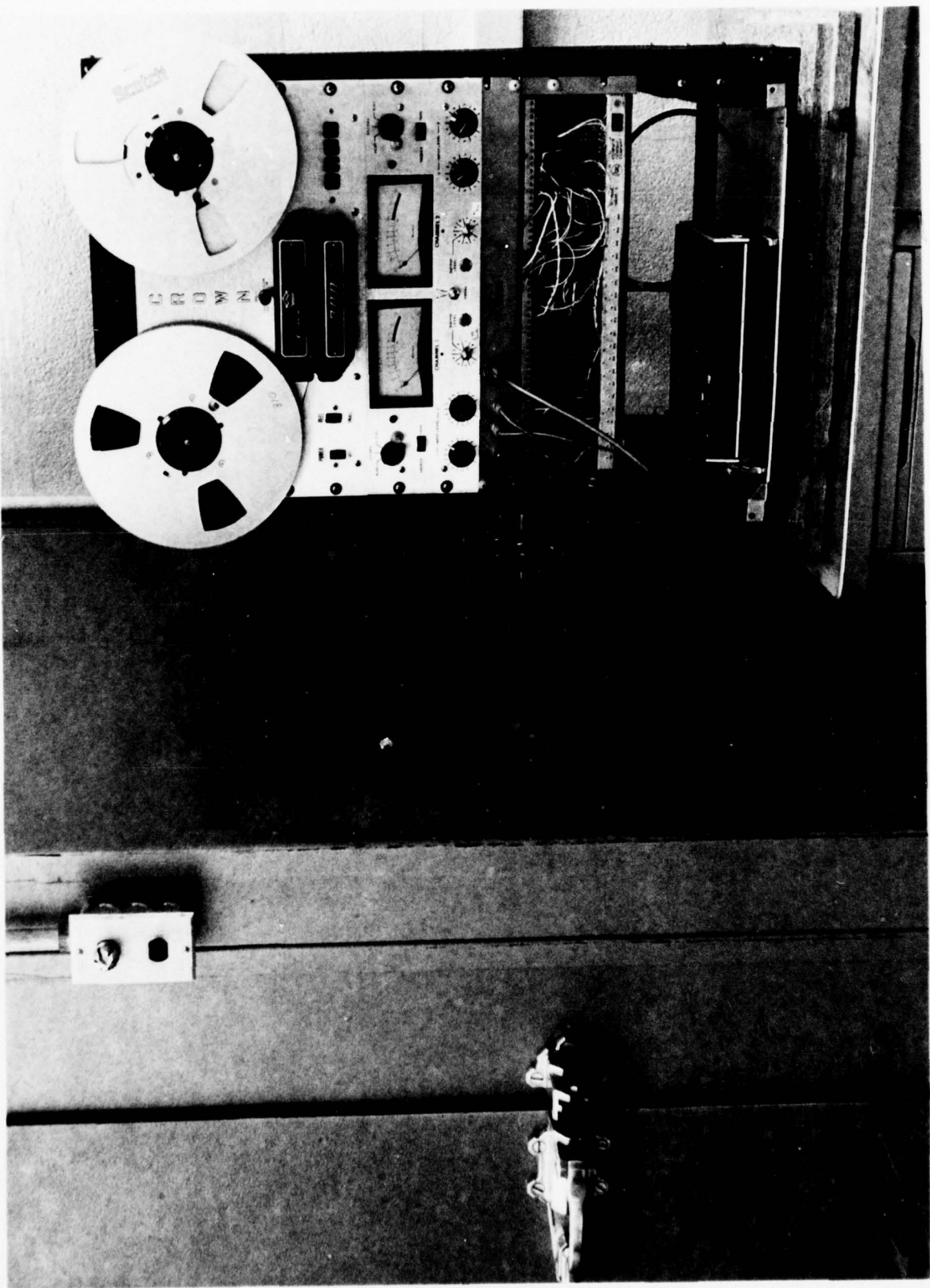
Figure 4.3 Photograph of the Equipment at the Test Location

State University.  For approval, a prospectus of the study was submitted.
This proposed approach was reviewed by a University Committee to ascertain
that no danger to the listeners would result.  Specifically, the levels
and durations of signals are such that no damage risk is incurred (American
Academy of Ophthalmology, 1957.)  Each subject was instructed as to his role
as participant in the study, and it was pointed out that under no conditions
would the sounds be so loud as to cause discomfort.

In order to eliminate the effect of shift in subjective loudness as
the probe signal-to-interfering-noise ratio changed, a balanced mixer
followed by an automatic loudness control was used.  The loudness as presented
to the subject was carefully maintained at a loudness level of 65 phons (GD)
(ISO R532.)  This loudness level was verified from 1/3rd octave band
measurements of the voltage function to the head phones and taking into account
the factory provided earphone calibration with the MX41/AR ear cushions.
This loudness level applied to the exposure signals as well as the probe
signal in noise.

The SNR was controlled by a balanced mixer with step increments of about
1/2 dB SNR per step.  The shaft of this mixer is connected to a potentiometer
which is used to vary the voltage of a voltage controlled oscillator.  The
oscillator frequency as a function of SNR of the output of the mixer was
calibrated at intervals of about 1/2 dB and it was this calibration data which
was used in the subsequent data reduction steps.  The combined frequency
uncertainty due to flutter and wow in the Crown tape player, the Sony cassette
recorder, and the final cassette player was less than $\pm$ 1/4th of the
frequency shift resulting from a change of SNR of 1/2 dB.

Each of the two exposure signals was analyzed and plotted against the
interfering noise background using 1/3rd octave analysis accurate to 1/2 dB.
This calibration was performed with a predetermined setting of the SNR mixer
setting (0).

## CHAPTER V

### EXPERIMENTS IN AUDITORY PATTERN DISCRIMINATION

#### 5.1 Introduction

A total of seven trained university students were used as listeners in experimental sessions spanning about six months. These students were concurrently, or had been, participants in other psychoacoustic studies at The Pennsylvania State University. There were three advanced standing female students and four graduate school males with normal hearing as verified by current audiograms. These listeners were free to administer test sessions on their own at their convenience with the following restrictions:

a) No two test sessions could run sequentially.

b) No more than two test sessions could be held on one day.

For all results presented in this chapter, the listeners used the modified threshold procedure. Also, they were instructed during the initial interviews and by voice comments on the tapes to:

"Indicate your choice as to which signal is being presented in the noise only when reasonably certain of your decision." The listeners were paid a fixed amount for each session they participated in. No feedback as to the number of correct classifications was given unless early results indicated either a tendency to guess at a very low SNR or a too conservative criterion which resulted in the probability of correct decisions being obviously higher than that obtained by other subjects. The results of these early results have been discounted from the data presented.

In the analysis of the results which follows, the term treatment is used to indicate a particular exposure set/probe combination. Hence, a situation where the signal pattern $\omega_1$ is associated with exposure signal A, signal pattern $\omega_2$ with exposure signal B, and signal pattern $\omega_1$ is used as the probe signal in an ocean ambient noise, is a specific treatment. The results of several treatments may, however, be combined (pooled) in some of the presentations. The statistical analysis consists of testing the effect of a treatment applied to the population consisting of the seven listeners. Except where specifically noted, no distinction is made between the subjects comprising the population. This method of presentation is tantamount to assuming that there are no individual differences (Murdock, 1968.)

Whenever data from magnetic tapes was used as input, a whitening filter was used to make the background noise nearly a broadband white noise. This whitening was used to reduce the overall signal dynamic range and is consistent with operational practice in sonar equipments.

## 5.2 Effects of High Pass Filtering

In studies of the intelligibility of speech in noise, it is found that a correspondence can be established between the cutoff frequency and listener performance (Pollack, 1948.) There seems to be a definite relationship between performance in various audiometric tests and the trained speech recognition process (Stevens and House, 1972.) For this reason, one of the principal areas of investigation centered around the effect of high pass filtering on the classification performance against marine sounds.

In this study there were 16 treatments involving four magnetic recordings of surface shipping. In all cases the exposure set consisted of such a recording and an identical signal which was high pass filtered with a sharp filter with 3 dB down point at 707 Hz. The frequency was chosen to correspond to an octave band edge for octave 30 complying with USA Standard S1.6-1967. Table 5.1 summarizes the treatments. In table 5.2 the significant aural characteristics of the source is given. It can be seen that the four ship sounds represent a considerable range in characteristics.

On the basis of considerations presented in a later section, the data for treatments differing only in the order of the exposure signals were pooled. The difference in 1/3rd octave spectral levels between the interfering noise and the probe signal is shown in Figures 5.1 and 5.2. The probe signal in each case is shown at a relative level corresponding to the point estimate of SNR required for listener response.

It can be seen from these data that the difference in SNR to respond to probe signals $\omega..$ and $\omega.._{HP}$ is in all cases less than 1 dB. This difference is not statistically significant when tested by the t test for difference of means. We may, therefore, pool all treatments which involve the same exposure signals in any order.

When the data are thus pooled, we can compute the probabilities $P_{ij}(C)$, the probability of a correct classification for pooled treatments, and $P_{ij}(E)$, the probability

May 28, 1975
CPJ:clb:cjg

| TREATMENT | A | B | PROBE | NO. OF EVENTS |
|---|---|---|---|---|
| I. 1 | $\omega_{10}$ | $\omega_{10HP}$* | $\omega_{10}$ | 12 |
| I. 2 | $\omega_{10}$ | $\omega_{10HP}$ | $\omega_{10HP}$ | 32 |
| I. 3 | $\omega_{10HP}$ | $\omega_{10}$ | $\omega_{10}$ | 16 |
| I. 4 | $\omega_{10HP}$ | $\omega_{10}$ | $\omega_{10HP}$ | 12 |
| I. 5 | $\omega_{11}$ | $\omega_{11HP}$ | $\omega_{11}$ | 4 |
| I. 6 | $\omega_{11}$ | $\omega_{11HP}$ | $\omega_{11HP}$ | 8 |
| I. 7 | $\omega_{11HP}$ | $\omega_{11}$ | $\omega_{11}$ | 28 |
| I. 8 | $\omega_{11HP}$ | $\omega_{11}$ | $\omega_{11HP}$ | 15 |
| I. 9 | $\omega_{12}$ | $\omega_{12HP}$ | $\omega_{12}$ | 18 |
| I. 10 | $\omega_{12}$ | $\omega_{12HP}$ | $\omega_{12HP}$ | 21 |
| I. 11 | $\omega_{12HP}$ | $\omega_{12}$ | $\omega_{12}$ | 12 |
| I. 12 | $\omega_{12HP}$ | $\omega_{12}$ | $\omega_{12HP}$ | 36 |
| I. 13 | $\omega_{14}$ | $\omega_{14HP}$ | $\omega_{14}$ | 6 |
| I. 14 | $\omega_{14}$ | $\omega_{14HP}$ | $\omega_{14HP}$ | 13 |
| I. 15 | $\omega_{14HP}$ | $\omega_{14}$ | $\omega_{14}$ | 8 |
| I. 16 | $\omega_{14HP}$ | $\omega_{14}$ | $\omega_{14HP}$ | 2 |

* $\omega \ldots_{HP}$ is signal pattern $\omega \ldots$ high pass filtered with a filter with cutoff frequency of 707 Hz.

Table 5.1. Summary of Treatments for Studies
of the Effects of High Pass Filtering

| SOUND PATTERN | DESCRIPTION OF ITS SOUND |
|---|---|
| $\omega_{10}$ | High speed rhythmic sound with pronounced bursts of high frequencies but of short duration. Sound is best described as that of a number of horses galloping across a concrete bridge. |
| $\omega_{11}$ | Very regular rhythm like that of a steam engine or slow train as heard inside. Some irregular high frequency popping. |
| $\omega_{12}$ | Noise-like without any noticeable rhythm. Sound like that of high speed air but at the same time having a low rumbling component. |
| $\omega_{14}$ | Noise-like with some irregular high frequency sounds like water splashing. Also, has a very pronounced steady tone in octave band 34 around $F_7$ . A hint of other tones which wax and wane. |

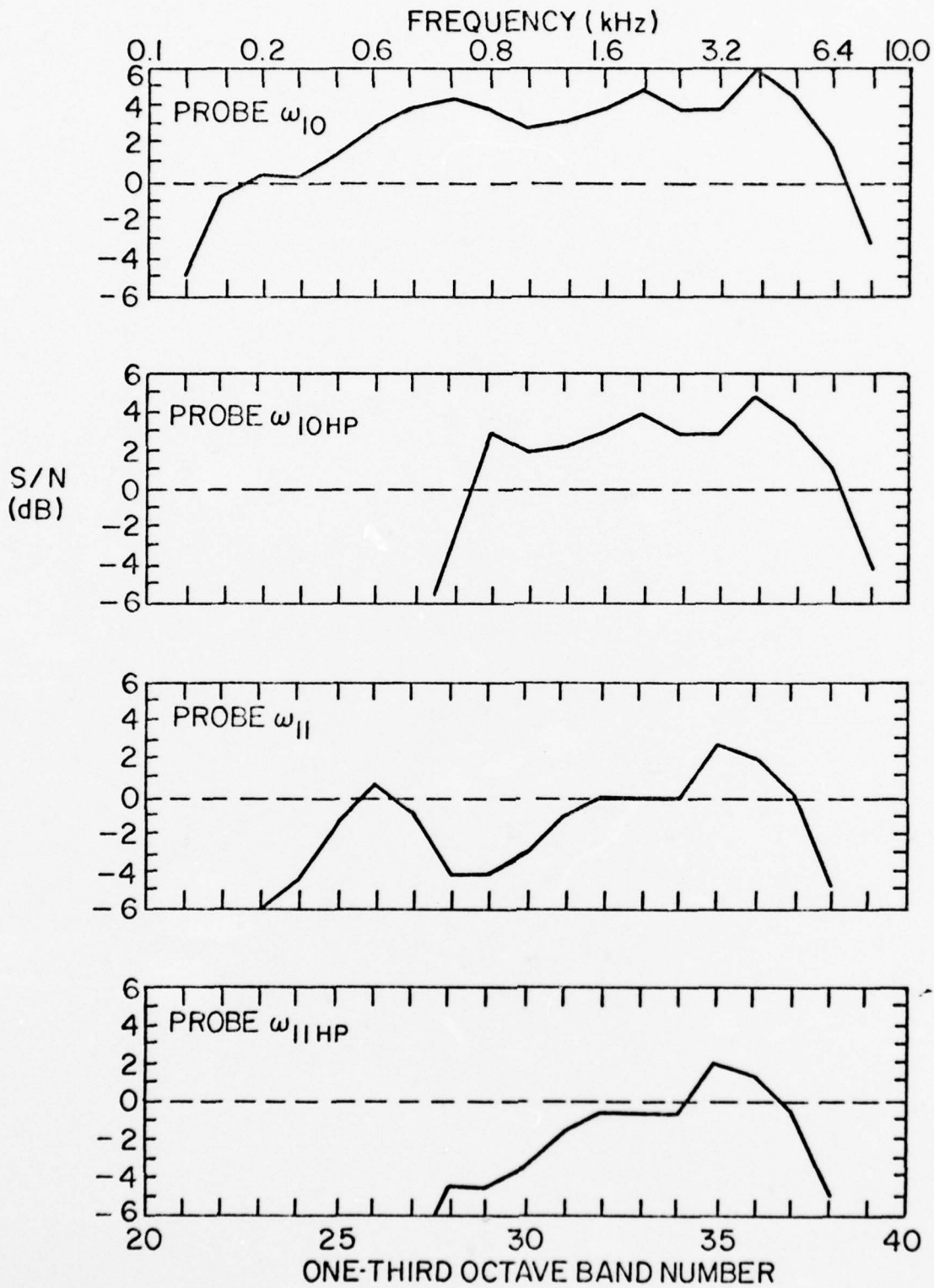Table 5.2. Summary of Sounds Associated with Patterns $\omega_{10}$, $\omega_{11}$, $\omega_{12}$ and $\omega_{14}$

Figure 5.1 Signal Excess at the Terminal Decision, Signal
Patterns $\omega_{10}$, $\omega_{10HP}$, $\omega_{11}$, $\omega_{11HP}$
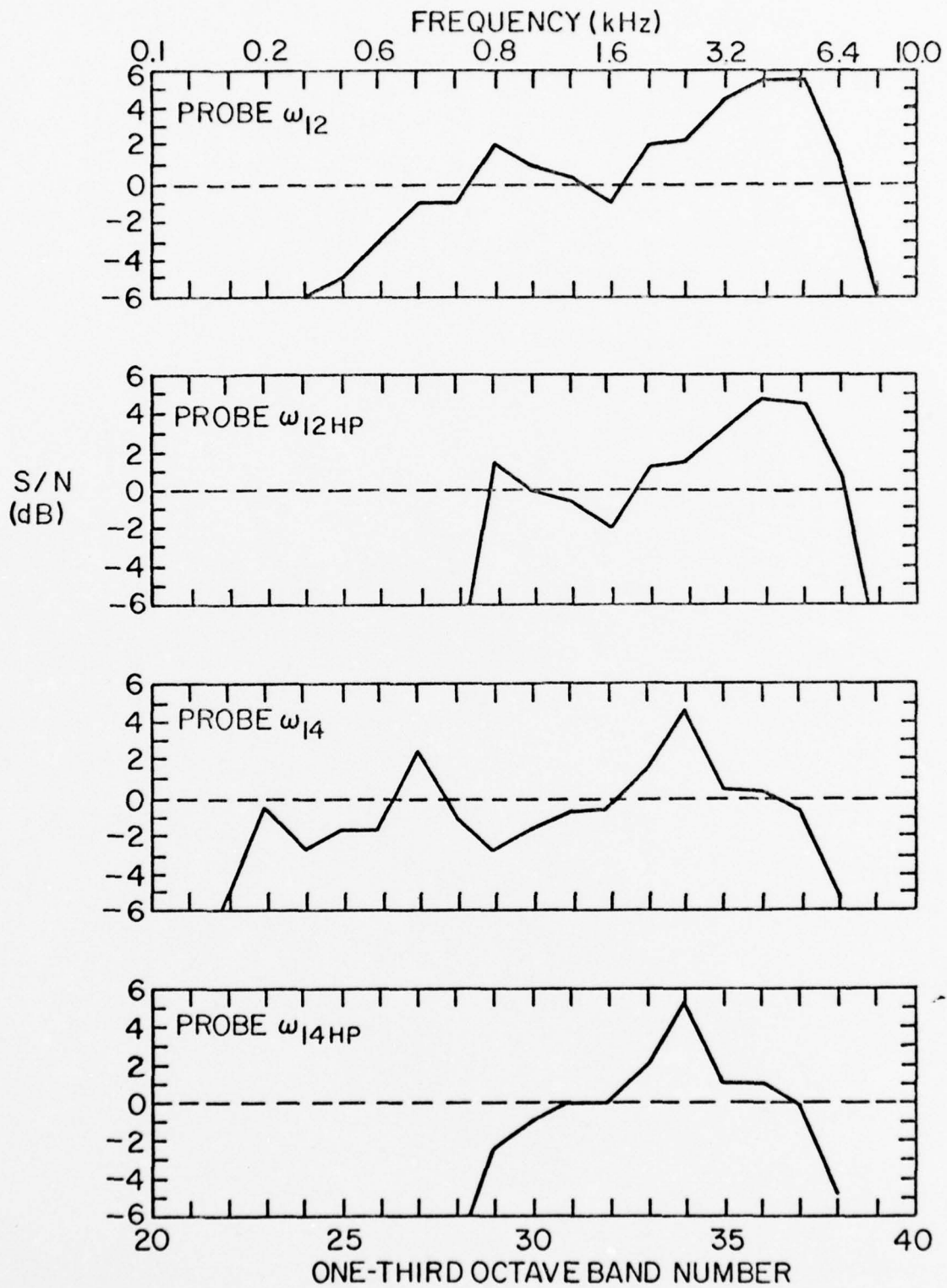
Figure 5.2  Signal Excess at the Terminal Decision, Signal
Patterns  $\omega_{12}$, $\omega_{12HP}$, $\omega_{14}$, $\omega_{14HP}$

of an incorrect classification for the pooled data. The number of no-response situations was small for this group of data and any events which did not end in a terminal decision for any subject of session were eliminated from consideration in any session by any subject. This handling of the data was deemed adequate even though it did reduce the total number of events useful in the analysis. Had the number of no-responses been greater, the situation would have been handled differently by the application of statistical concepts valid for marginal distributions (Winer, 1962.)

The results for the pooled data are shown in Table 5.3. These are the point estimates for the SNR required for a terminal decision under the criterion of "reasonably certain." With each set of exposure signals $\omega..$ and $\omega.._{HP}$ there is associated an observed probability of correct decision. This too is a point estimated of the underlying probability governing the likelihood of a specific outcome to the experiment each time it is performed (Swets, 1964.) Assuming that the pooled events comprise Bernoulli trials with equal probability of occurrence, the binomial distribution

$$f(x) = \binom{n}{X} p^{X}(1 - p)^{n-x} \; , \; x = 0,1,2 \ldots n$$
$$= 0 \text{ elsewhere}$$

applies for $x$ observed correct classifications in $n$ trials (Hogg and Craig, 1970.) Again this approach assumes no individual differences between listeners, and no order effects, and no differences attributable to the probe used. Murdock (1968) shows that this binomial distribution may be transformed to an approximately Gaussian distribution using the transformation

$$\theta = 2 \text{ arc sin } \sqrt{X/n} \quad .$$

The variance of the transformed variable will be:

$$\sigma_{\theta}^{2} = 1/n \quad .$$

Hence, $\theta$ is $n(\theta_{o}, 1/n)$ where $\theta_{o}$ is the transformed sample probability of correct classification.

| PATTERNS | MEAN SNR $\overline{SNR}$ | STANDARD DEVIATION $S_{SNR}$ | OBSERVED $P(c)$ | 90 PER CENT CONFIDENCE INTERVAL | |
|---|---|---|---|---|---|
| | | | | ON $\overline{SNR}$ | ON $P(c)$ |
| $\omega_{10}$, 10HP | 2.99 db | 2.81 db | 0.67 | 2.35, 3.63 | 0.62, 0.72 |
| $\omega_{11}$, 11HP | 2.41 db | 3.48 db | 0.80 | 1.49, 3.33 | 0.74, 0.85 |
| $\omega_{12}$, 12HP | 4.80 db | 3.20 db | 0.59 | 4.15, 5.45 | 0.55, 0.62 |
| $\omega_{14}$, 14HP | 4.92 db | 3.81 db | 0.70 | 3.58, 6.26 | 0.62, 0.77 |

Table 5.3.  Summary of Results Pooled Across
Order and Probe

The 90 per cent confidence interval on the mean SNR when the sample standard deviation is used is:

$$[\overline{SNR} - t(0.95,\nu) \, {}^{S}/\sqrt{N} \; < \mu_{SNR} < \overline{SNR} + t(0.95,\nu) \, {}^{S}/\sqrt{N} \, ]$$

where $t(0.95,\nu)$ is the single tail t statistic at the 95 per cent level with $\nu$ degrees of freedom (Weinberg and Schumaker, 1962.)  In the limit of a large number of degrees of freedom

$$t(0.95,\nu) = 1.69 \quad , \; \nu \geq 33.$$

The 90 per cent confidence interval on the probability of correct classification is given approximately by:

$$[\theta_o - Z(0.95) \, {}^{1}/\sqrt{n} \; < \mu_o < \theta_o + Z(0.95) \, {}^{1}/\sqrt{n} \, ]$$

where $Z(0.95)$ is the Z score corresponding to the one tailed normal distribution (Murdock, 1968.)  Also,

$$Z(0.95) = 1.69.$$

These confidence intervals are also given in Table 5.3.  Note that the confidence interval on the probability of a correct decision is quite large ($\sim \pm 5$ per cent) even when a large number of events are pooled.  Figure 5.3 shows the behavior of this confidence interval for the two cases where the actual probabilities of the occurrence of a correct decision are 0.7 and 0.5 respectively.  It is seen that the number of events must exceed about seventy for the 90 per cent confidence interval to be within $\pm 10$ per cent of this value.  The number of events needed to estimate this parameter is therefore greater than 50 and preferably 70 if the confidence interval is to be at all useful.  Workers using the forced response techniques routinely use 300 or more events to estimate each data point (Swets, 1964, Green and Swets, 1966, Green, 1960a)

Another product of this research is the observation of the types of errors made by the listener.  Table 5.4 summarizes these results.  It is seen that the types of misclassifications are not strongly biased, and these data tend to support the conclusion that the listener is equally likely to misclassify $\omega..$ as $\omega.._{HP}$ as he is to misclassify $\omega.._{HP}$ as $\omega..$ .
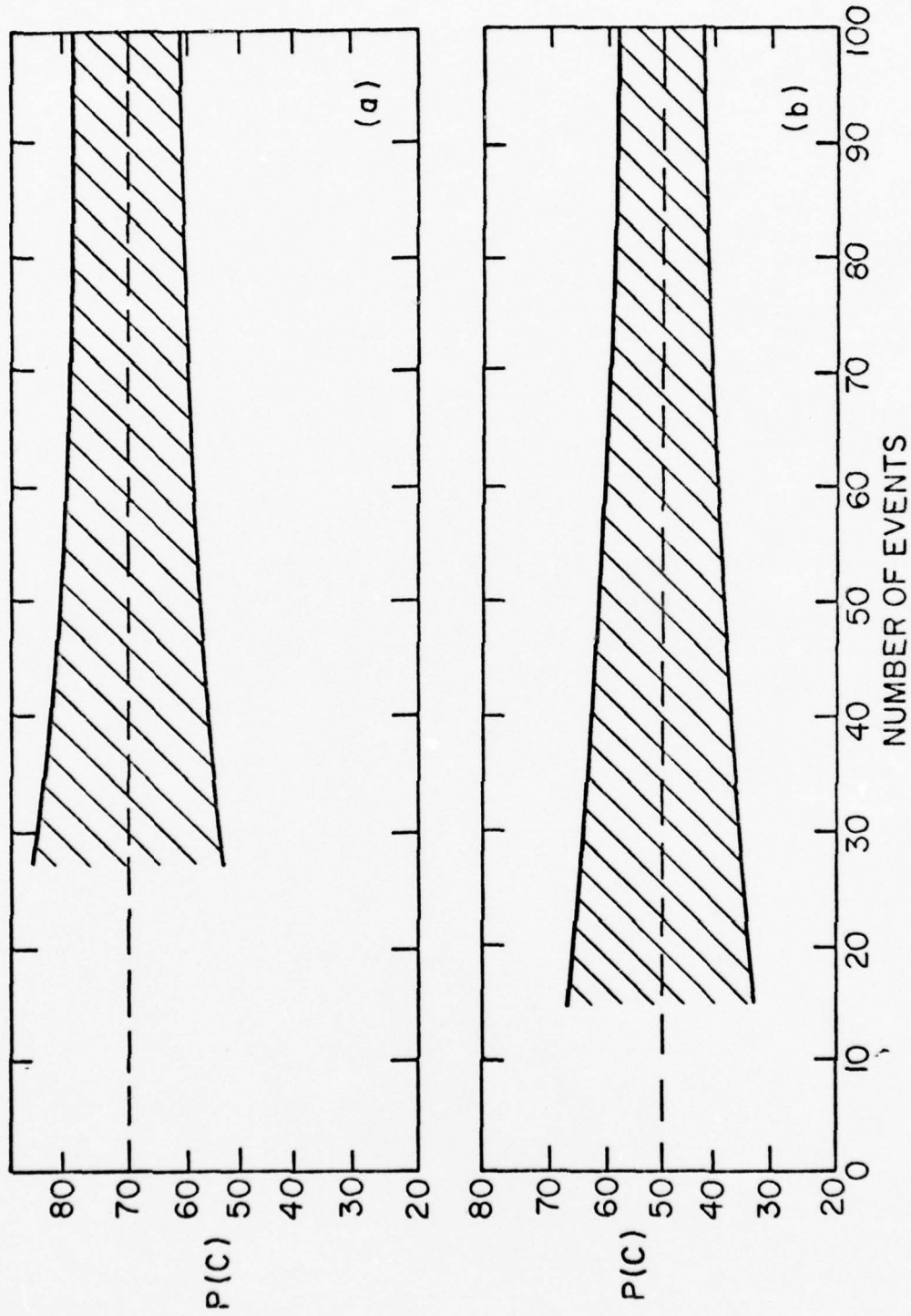
Figure 5.3  Confidence Intervals on P(c)  vs. Number of Events,
(a) For P(c) = 0.7,   (b) For P(c) = 0.5

| PROBABILITY EXPRESSION | OBSERVED | PROBABILITIES |
| | $\alpha$ | $\beta$ |
| --- | --- | --- |
| $P(\hat{\omega}_{10} \mid \omega_{10HP})$ | 0.19 | |
| $P(\hat{\omega}_{10HP} \mid \omega_{10})$ | | 0.14 |
| $P(\hat{\omega}_{11} \mid \omega_{11HP})$ | 0.07 | |
| $P(\hat{\omega}_{11HP} \mid \omega_{11})$ | | 0.13 |
| $P(\hat{\omega}_{12} \mid \omega_{12HP})$ | 0.18 | |
| $P(\hat{\omega}_{12HP} \mid \omega_{12})$ | | 0.22 |
| $P(\hat{\omega}_{14} \mid \omega_{14HP})$ | 0.17 | |
| $P(\hat{\omega}_{14HP} \mid \omega_{14})$ | | 0.13 |

Table 5.4.  Misclassification Probabilities
for High Pass Filtering Experiments

## 5.3 Temporal and Tone Effects

Again referring to the findings from speech intelligibility, the intensity, frequency, and time structure of a signal would be expected to be important (Licklider and Miller, 1951). Two sets of experiments were performed to address this question. The first of these compared the classification performance when the exposure set consisted of two signals which were spectrally similar but differed in the signal envelope structure. The second set of experiments investigated the effect of "soft" limiting of the signal.

A total of five treatments were tested in which the patterns were a recording of a marine sound (pattern class $\omega_{..}$) and a spectrally similar shaped Gaussian noise (pattern class $\omega_{..SN}$). These treatments are summarized in Table 5.4. The shaped noise was obtained by passing the output of a broadband noise generator through a General Radio multi-filter with the 1/3rd octave weights adjusted to give a resulting spectrum which by 1/3rd octave analysis differed by no more than 1 db in any band from the recorded marine source spectral level.

The shaped noise differed from the recorded source both in its amplitude vs. time behavior and in its narrowband spectral structure. Sound pattern $\omega_{10}$ is characterized by a pronounced envelope modulation but has a relatively smooth averaged narrowband spectrum. In contrast, sound pattern $\omega_{14}$ has little systematic variation of level with time but has a tonal quality due to strong narrowband components. When the data are pooled across order and probe, the resulting SNR and P(C) are as shown in Table 5.6.

To test the effect of limiting on signals, the tape recorded signal was passed through a non-linear circuit with the transfer characteristic shown in Figure 5.5. The input level was so adjusted that limiting began at the $1\sigma$ point of the time waveform. Such a transfer function, while modifying the detailed amplitude vs. time structure of the signal, does not greatly reduce the intelligibility of speech (Licklider, Bindra and Pollack, 1948).

Five treatments using two marine source recordings were investigated. Of the 55 events, only 18 resulted in a terminal decision and of these 55 per cent were correct classifications. The treatments and number of events tested in each case are summarized in Table 5.5. The responses and listener comments

| TREATMENT | A | B | PROBE | NO. OF EVENTS | ANALYSIS OBJECTIVES |
|---|---|---|---|---|---|
| II. 1 | $\omega_{10}$ | $\omega_{10SN}$ | $\omega_{10SN}$ | 5 | To investigate the effect of replacing a recorded source by a spectrally similar shaped noise ($\omega \ldots _{SN}$). |
| II. 2 | $\omega_{10SN}$ | $\omega_{10}$ | $\omega_{10}$ | 18 | |
| II. 3 | $\omega_{14}$ | $\omega_{14SN}$ | $\omega_{14}$ | 11 | |
| II. 4 | $\omega_{14}$ | $\omega_{14SN}$ | $\omega_{14SN}$ | 17 | |
| II. 5 | $\omega_{14SN}$ | $\omega_{14}$ | $\omega_{14}$ | 17 | |
| III. 1 | $\omega_{10}$ | $\omega_{101\sigma}$ | $\omega_{10}$ | 11* | To ascertain if $1\sigma$ limiting of a recorded source ($\omega \ldots _{1\sigma}$) has a discernable effect on the signal. |
| III. 2 | $\omega_{10}$ | $\omega_{101\sigma}$ | $\omega_{101\sigma}$ | 11* | |
| III. 3 | $\omega_{14}$ | $\omega_{141\sigma}$ | $\omega_{14}$ | 10* | |
| III. 4 | $\omega_{14}$ | $\omega_{141\sigma}$ | $\omega_{141\sigma}$ | 11* | |
| III. 5 | $\omega_{141\sigma}$ | $\omega_{14}$ | $\omega_{141\sigma}$ | 10* | |

\* No. of events attempted.

Table 5.5.   Summary of Treatments Investigating
Temporal and Tone Effects

| PATTERNS | MEAN SNR $\overline{SNR}$ | STANDARD DEVIATION $S_{SNR}$ | OBSERVED P(C) | 90 PER CENT ON $\overline{SNR}$ | CONFIDENCE INT. ON P(C) |
|---|---|---|---|---|---|
| $\omega_{10}$, $\omega_{10SN}$ | -3.23 db | 2.15 db | 0.78 | -3.99, -2.47 | 0.61, 0.87 |
| $\omega_{14}$, $\omega_{14SN}$ | -0.45 db | 3.26 db | 0.84 | -1.27, 0.37 | 0.72, 0.92 |

Table 5.6.   Summary of Results for Tests with
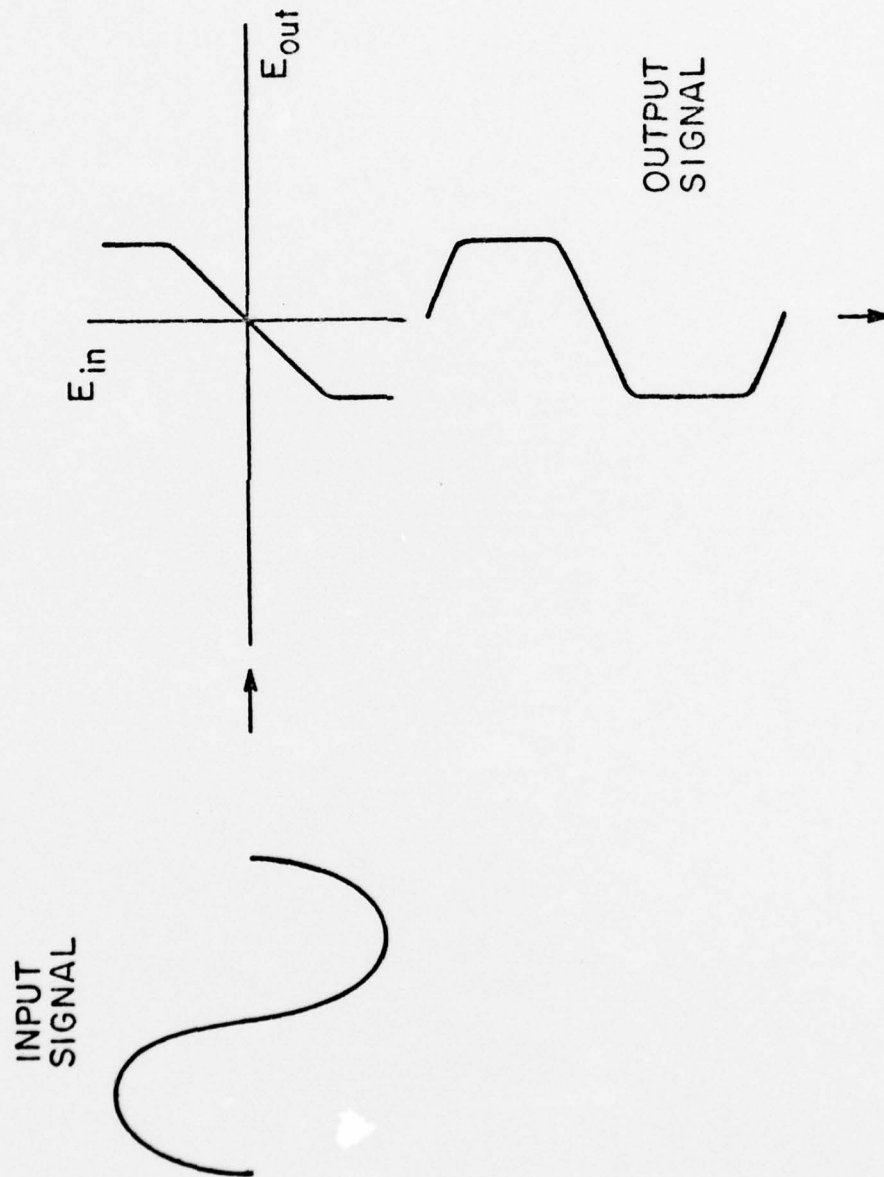Spectrally Shaped Noise

Figure 5.4  Soft Limiting Circuit Time Domain Transfer Function

indicate that there is little subjective difference in the sounds attributable to the limiting process. Those responses which were made occurred at a SNR near the maximum presented. At this high SNR the signals are little contaminated by the background noise and the task is one of matching an essentially pure probe to the exposed signal set.

## 5.4 Learning Effects and Experimental Biases

In any experimental technique such as the modified threshold procedure used here, the question of biases introduced by that technique must be investigated. Referring to Figure 4.1 it is seen that in every case signal B is presented just prior to the response period. It is worthwhile, therefore, to ask the question.

Does the order of the signal exposures significantly effect the experimental outcome?

That is, can we disprove the null hypothesis that there is no order effect due to the sequence of the exposure signals? From Tables 5.1 and 5.4 we see that the below-listed pairs of treatments differ only in the order of the exposure signal.

> I. 1 and I. 3, I. 2 and I. 4, I. 5 and I. 7,
> I. 6 and I. 8, I. 9 and I.11, I.10 and I.12,
> I.13 and I.15, I.14 and I.16, II.3 and II.5.

When tested by applying the t test for difference of means $(\overline{SNR})$, only in the case of treatment pairs I.1 and I.3 is the null hypothesis disproved at the 0.1 level. Some doubts about the calibrations in this case may account for this discrepancy. On the basis of these results, it appears that there is not a significant order effect. These findings allow pooling of the data without regard to exposure order.

Another consideration in this type of experiment is that of the time factor. In studies of auditory memory span, it is routine to present an exposure signal followed by a shadowing stimulus prior to the probe presentation (Parkinson, Parks and Kroll, 1971). The effect of the shadowing stimulus is quite pronounced when tests are done with isolated letters on phonemes in a non-rehearsal situation.

If we now consider the low SNR presentation at the beginning of the response period as a shadowing stimulus, we may expect the performance to decrease as the time to respond increases. Also, there is the question of memory latency which may enter in.

In order to test for this effect, the rank correlation coefficient of SNR to respond vs. the time to respond was computed. This value varied between 0.03 to 0.33. With these data alone, however, it is difficult to determine if the rank correlation coefficients on the order of 0.3 does in fact indicate a significant correlation. If, for example, the initial SNR were always to be fixed, the responses with higher SNR would be perfectly correlated with response time by virtue of the fact that the SNR varies in a predetermined fashion with time. We have purposefully randomized this initial SNR to reduce this effect. It is not immediately clear how much remaining correlation one should expect. In a later chapter, a revised experimental approach is suggested which may allow an answering of the question:

How much effect is there due to contamination of the reference sounds in memory due to time lapse or shadowing?

The responses for two pairs of signal patterns are plotted in Figures 5.5 and 5.6. It is seen that there may in fact be some tendancy for the required SNR to increase with response time. The degree of this trend, if it does in fact exist, cannot be ascertained. It should also be pointed out that the present set of data does not conform to the test without rehearsal done by other workers. The listeners were in fact encouraged to make notes which could help them in the classification task.
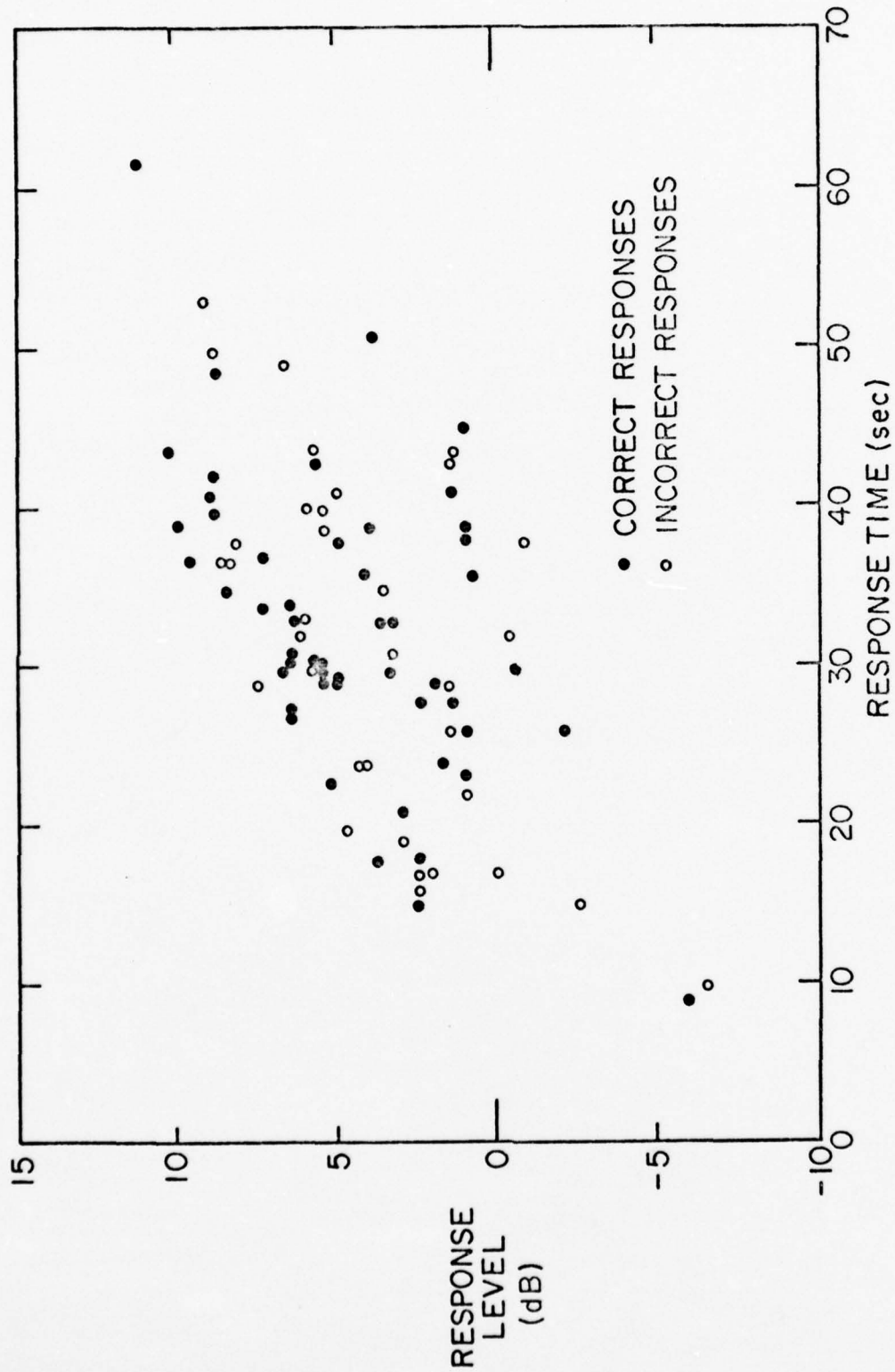
Figure 5.5  Responses as a Function of Time into the Response Period, Signal Pair $\omega_{12}$, $\omega_{12HP}$
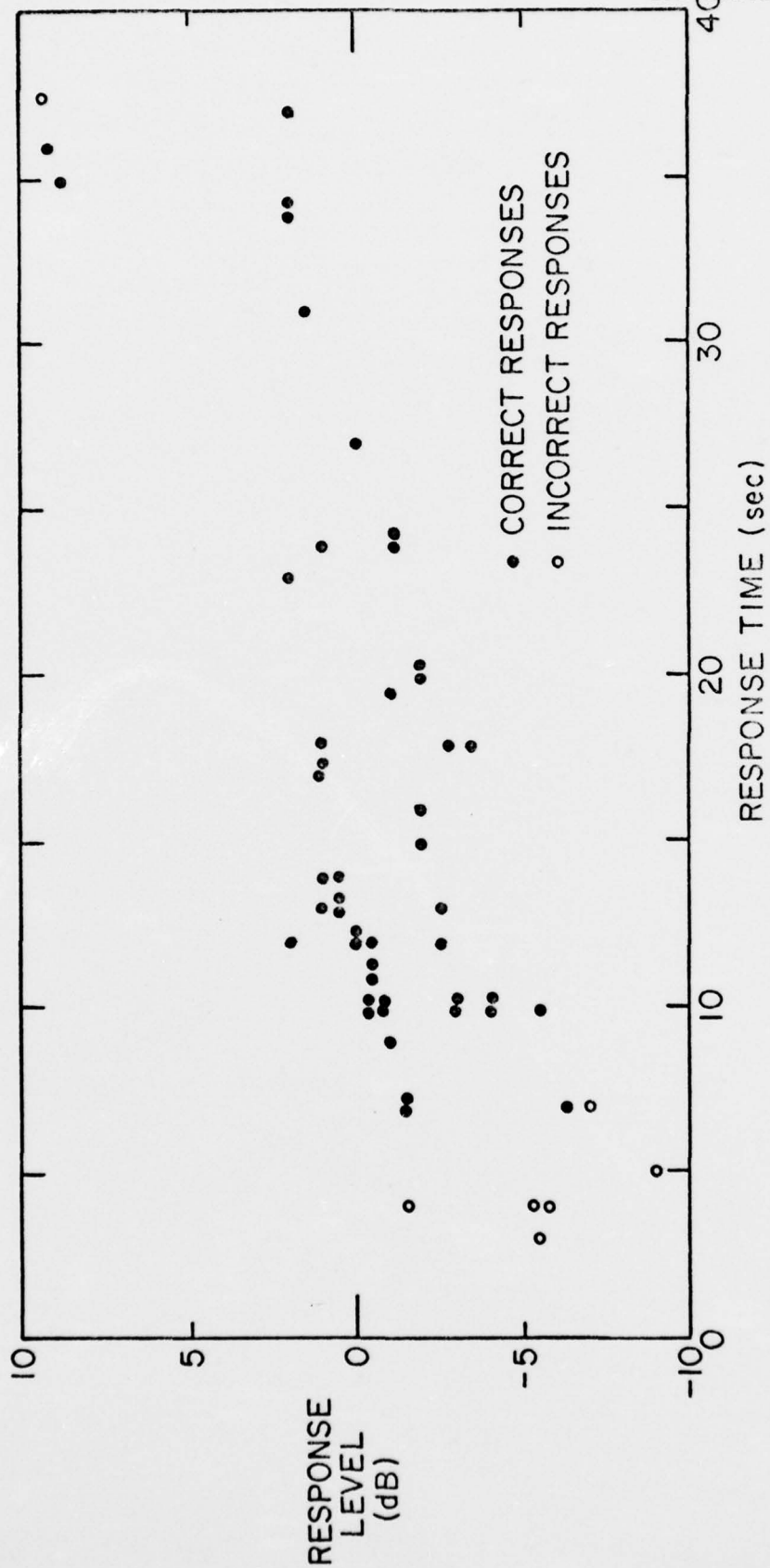
Figure 5.6  Responses as a Function of Time into the Response Period, Signal
Pair $\omega_{14}$, $\omega_{14HP}$

CHAPTER VI

DISCUSSION OF THE EXPERIMENTAL RESULTS

## 6.1 Interpretation of the Results from the Point of View of Signal Detection Theory

In this chapter we will analyze the results of the experiments to date from the point of view of signal detection theory. Specifically, the questions raised in Chapter II about the dichotomous feature classification problem will be pursued.

Consider an auditory recognition task where two signal patterns $\omega_i$ and $\omega_j$ differ only by the presence of feature $\Omega_k$ in one case. This feature will have associated with it a detectability d'. We will then model the classification task as a hypothesis test designed to determine the presence of feature $\Omega_k$. It is certainly not obvious that the ability simply to detect the presence of a feature is adequate to make the pattern classification. There are, after all, other features present in both signal patterns which can perhaps confound the classification task. It will be seen, however, that this simple approach does yield some rather consistent results.

## 6.2 Analysis of Results for High Pass Filtering

When the two pattern classes differ in that one ($\omega..._{HP}$) has the low frequency noise components eliminated by filtering, we can define a dichotomous feature $\Omega_{LP}$ which embodies all of the low frequency information. This is itself a feature complex in that it consists of at least the following components:

a) A band-limited noise with non-uniform spectral density in the auditory band from about 120 Hz to 707 Hz. (Limited by the tape player at the low end.)

b) Amplitude modulation components of the above band-limited noise

c) Perhaps some time dependent shifts in frequency of the spectral components of this band-limited noise.

d) Tonals falling in this band of frequencies.

Further assume that this dichotomous feature $\Omega_{LP}$ can be characterized as a first approximation as a band of noise with rectangular bandwidth $W_{eff}$ and corresponding spectral level $S_{eff}$. The classification under these assumptions then becomes:

Conclude $H_o$; decide signal pattern is $\omega..._{HP}$

Conclude $H_1$; decide signal pattern is $\omega..$.

where the hypotheses are:

$H_o$:  Feature $\Omega_{LP}$ is absent

$H_1$:  Feature $\Omega_{LP}$ is present.

The problem of detecting the band of noise was discussed in Chapter II. We saw there that the optimum detector performance is given by Equation 2.2.

Green (1960a) extensively investigated listener performance in the detection of bands of noise in noise problems. He found that $d'_{obs} = a\ d'_{opt}$, the observed detectability, for a large range of WT. For a $P(C) = 0.75$, the value of a was between 0.25 and 0.33 depending on subject. This finding was seen to hold for a large range of center frequencies ($f_c$) for the band of noise. The range of $f_c$ was from 400 to 6000 Hz, W varied from 655 to 5143 Hz, and T varied from 3 to 1000 msec. The results above 300 msec seemed to deviate from the expected behavior, however. Also, he measured an $f_c$ of 7500 Hz but found some anomolies which he attributes to the effect of earphone behavior.

Figure 6.1 combines all of Green's findings on probability paper. This figure includes the result of two experiments with five different subjects and with the following set of parameters.

$f_c$ = 400, 800, 1500, 2500, 4500, 6000 Hz.

W  = 655, 3862, 5143 Hz.

T  = 3, 10, 30, 100, 300 msec.

These results are for a 2AFC procedure and the data is normalized by using the equation:

$$10 \log d'_{opt} = 10 \log (WT)^{1/2}\ \frac{\sigma_S^2}{\sigma_n^2}\ [1/2(\sigma_S^2/\sigma_n^2)^2 + (\sigma_S^2 + \sigma_n^2) + 1]^{-1/2} . \quad (6.1)$$

In our results, the detectability of feature $\Omega.._{LP}$ can then be thought to correspond to the band of noise used in Green's experiments. When the equivalent square bandwidth $W_{eff}$ is computed for the four signals used, it is seen that most of the signal power associated with $\Omega.._{LP}$ falls in an octave band centered nominally about 500 Hz. Hence
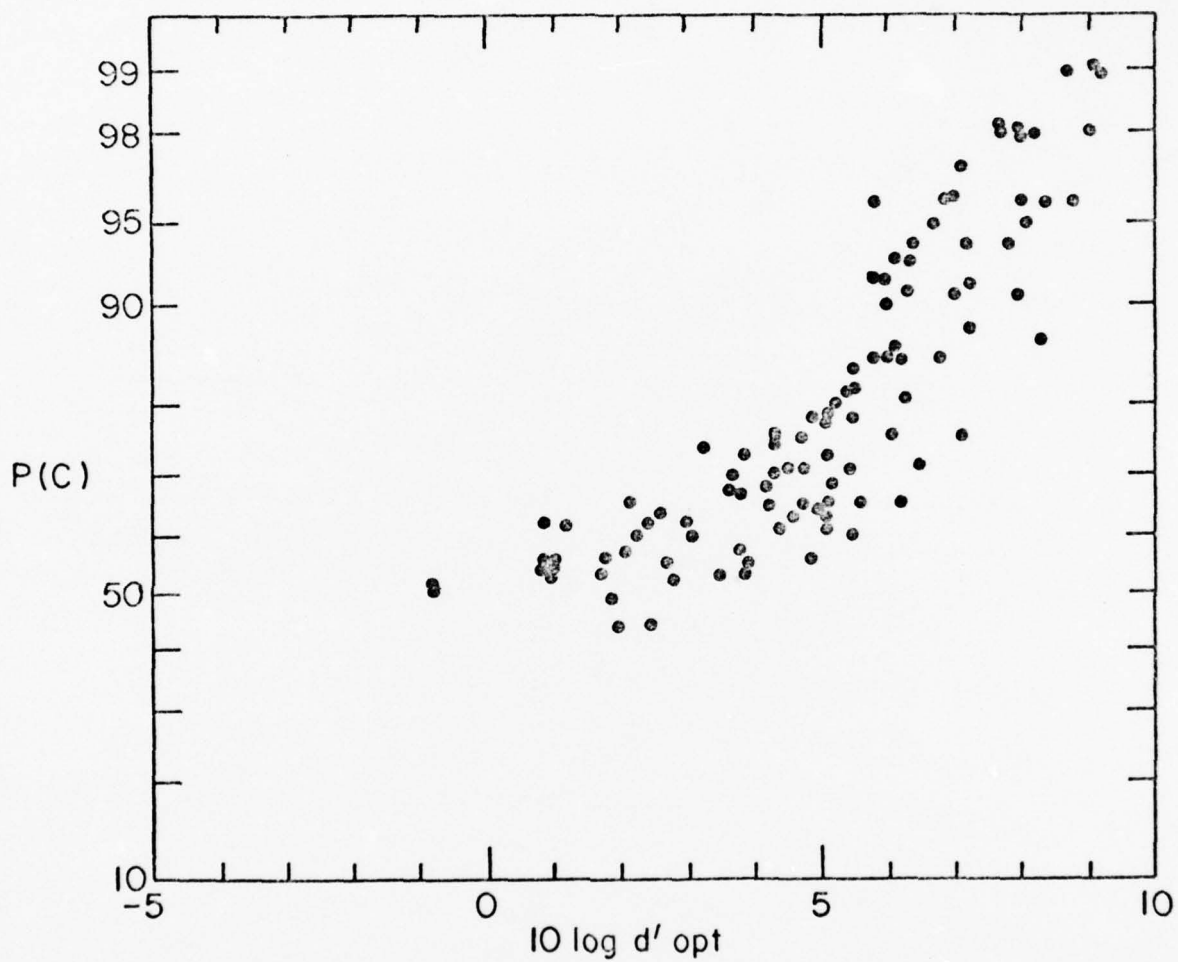
$$W_{eff} \cong 345 \text{ Hz}.$$

Figure 6.1   Summary of Green's Results for the Detection of a
Noise in Noise

The ratio between the signal and noise spectral level

$$S^2_{eff}/N^2 = \sigma^2_\Omega/\sigma^2_n$$

is shown in Table 6.1 for the four pairs of signal patterns used.

If we are to normalize our data as does Green, how do we define T, the system integration time, in the Equation 6.1? The length of time for which the SNR was maintained constant in the experiment was 2 seconds. However, the listener could be integrating over successive steps. Or, the time T may be much shorter due to some as yet unmentioned mechanism. We see that the expression for $d'_{opt}$ implies continuously increasing detectability with time. It is found, however, that a listener extracts all useful information from an acoustic stimulus in the first few hundred milliseconds (Tanner and Sorkin, 1972.) In fact, the behavior observed by Green for the data with duration of 1 second indicates that

$$d'_{obs} \text{ for } T = 1000 \text{ msec} \cong d'_{obs} \text{ for } T = 300 \text{ msec.}$$

A detailed analysis of the response behavior for the modified threshold procedure indicates that listeners exhibit comparable behavior in this type of experiment. The number of observed responses as a function of delay from the step change in SNR is plotted in Figure 6.2. This data is for 160 responses chosen at random from among the approximately 480 available events. In Figure 6.2 the independent variable $\tau$ is the delay to respond from the onset of the step change in SNR. The peak in the number of responses at 600 msec is significant at the 0.1 level under the assumption of a Poisson distribution for the number of responses per 100 msec of delay. This observation supports the view that each step change in SNR can be treated as a detection opportunity. After a delay, the listener has extracted all of the additional information provided by the change in detectability associated with the change in SNR and either makes a terminal decision or defers the decision until another such change occurs. It is noteworthy that this result is observed in spite of the fact that listeners were unable reliably to describe the way in which the SNR changed when asked to verbalize the auditory sensation. They were unable consciously to tell if the SNR changed continuously or in steps!

| FEATURE | $10 \log(\sigma_g^2/\sigma_n^2)$ | POINT ESTIMATE $10 \log d'_{mt}$ | *INTERVAL ESTIMATE $10 \log d'_{mt}$ | POINT ESTIMATE $P(c)$ | INTERVAL ESTIMATE $P(c)$ |
|---|---|---|---|---|---|
| $\Omega_{10LP}$ | 3.3 | 10.86 | 10.61, 11.10 | 0.67 | 0.62, 0.72 |
| $\Omega_{11LP}$ | 1.1 | 9.87 | 9.35, 10.32 | 0.80 | 0.74, 0.85 |
| $\Omega_{12LP}$ | -2.4 | 7.63 | 7.14, 8.1 | 0.59 | 0.55, 0.62 |
| $\Omega_{14LP}$ | 1.6 | 10.12 | 9.4 , 10.72 | 0.70 | 0.62, 0.77 |

* Interval Estimates are the 90 per cent confidence intervals.

Table 6.1. Results of Modified Threshold Procedure for Classification Under Dichotomous Features, High Pass Case
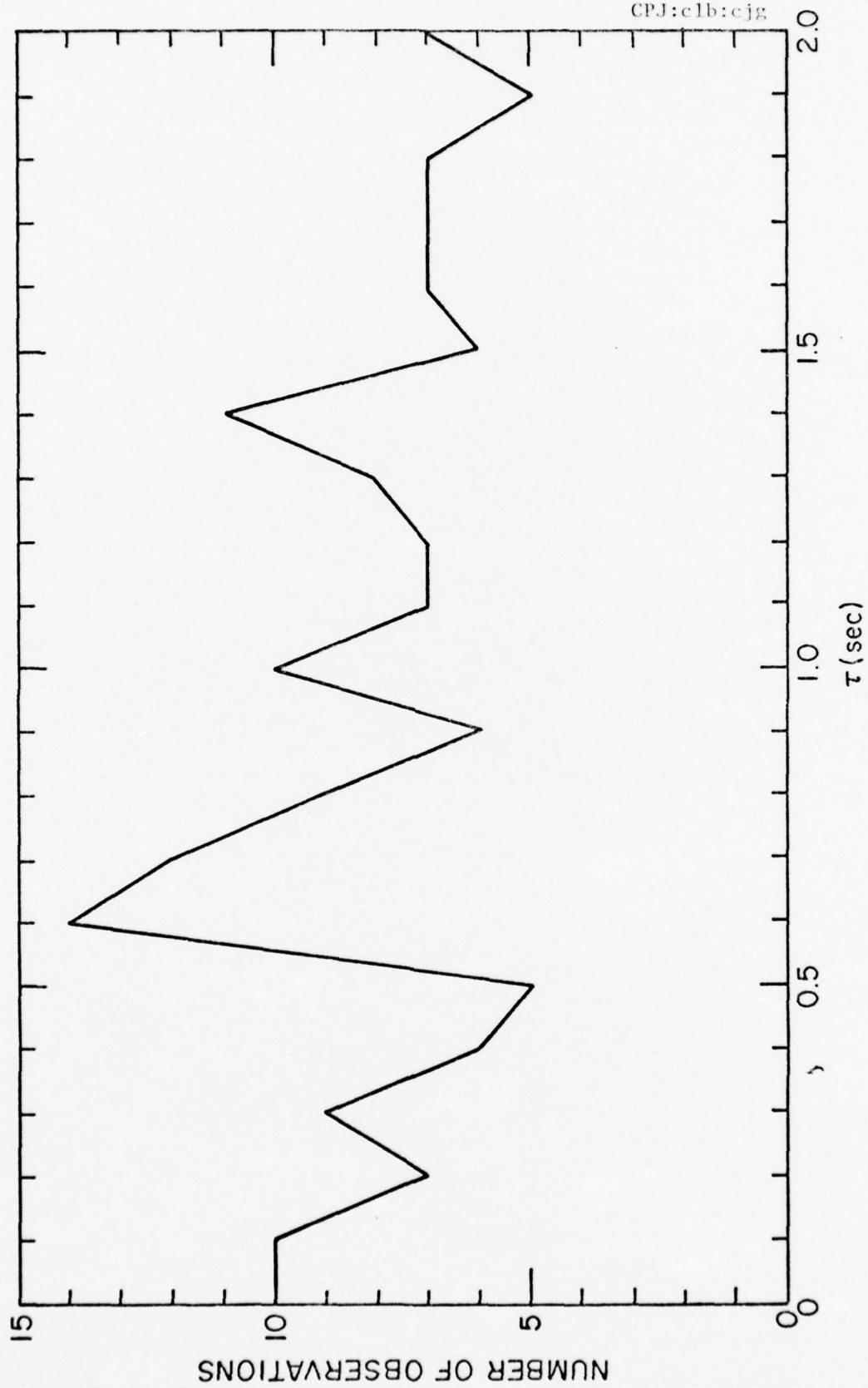
Figure 6.2  Distribution of Responses as a Function of Delay from
the Step Change in SNR

From these data, and allowing for a response time of 100 msec from the time of the terminal decision to the recording of that response, it was inferred that an integration time of 500 msec is reasonable.

We can then define a detectability

$$d'_{mt} = (W_{eff} T_{eff})^{1/2} \frac{\sigma^2_\Omega}{\sigma^2_n} [1/2(\sigma^2_\Omega/\sigma^2_n)^2 + (\sigma^2_\Omega + \sigma^2_n) + 1]^{-1/2}$$

where

$$W_{eff} = 354 \text{ Hz}, \quad T_{eff} = 0.5$$

and $d'_{mt}$ is the observed detectability using the modified threshold procedure. The point estimate and 90 per cent confidence interval on $10 \log d'_{mt}$ is also given in Table 6.1.

When $d'_{mt}$ is compared to $d'_{obs}$ from Green's results, we find that

$$10 \log d'_{mt} - 10 \log d'_{obs} = 5.3 \pm 1.2$$

when the point estimate of $P(C)$ is used for comparison purposes. It is seen that the four signals give consistent results one to another in spite of the fact that they differ significantly in their overall sound. Also, in spite of the fact that the listener instructions in all cases was to "....indicate your choice only when reasonably certain.", the probabilities of a correct response differed significantly especially in the case of signal pattern $\omega_{12}$. The reduced $P(C)$ was also made at a significantly lower SNR, however. The behavior is consistent with the results obtained by Green and conforms to the model. The reason for this difference in what is apparently the listener criterion is treated in more detail in a later section of this chapter.

What of the five dB difference between $d'_{mt}$ and Green's results? Up to this point we have treated the classification as if it were simply a detection problem on feature $\Omega_{LP}$. However, the listener, in order to establish a threshold for the terminal decision, must also listen to the rest of the signal. Stallard and Leslie, (1974) conclude on theoretical grounds that the difference between a 2AFC experiment and passive sonar detection performance should be about 5.4 dB. Their reasoning is as follows:

a) The effect of frequency uncertainty because two bands of noise must be attended to introduces some decrease in detectability.

b)  Time uncertainty about the onset time of the signal also
    has the same effect.

c)  They modeled the passive sonar problem as a YN task and
    added another 1.5 dB for the difference between the effi-
    ciency of a YN and 2AFC test.

This latter correction was made in our case by the definition of $d'_{opt}$
hence is not applicable.  It is, therefore, suggested by the analysis pre-
sented by Stallard and Leslie that the present results should be related to
Green's (1960a) results by a factor of about 4 dB, or:

$$d'_{mt} = 2.5 \cdot d'_{obs} \quad , \text{ theoretically.}$$

The remaining discrepancy of about 1 dB is of doubtful significance.
More important is the matter of which SNR to use in computing the $d'_{mt}$ at
the time of the terminal decision.  If the listener somehow integrates over
previous observations, the apparent SNR at the t'th observation would be
higher than measured.  That is, if at observation t-1 the listener is aware
of the fact that the log likelihood ratio almost exceeded the threshold,
this is useful information and, hence decreases the uncertainty.  (Raisbeck,
1963.)  In fact, Swets and Green (Swets, 1964) have demonstrated that a
listener is indeed capable of integrating the information in successive ob-
servations in very specialized circumstances.  In general, however, they
note that:

"This analysis leaves little doubt that the assumption of no integration
over successive observations is a good one. . . . ."

That is not to say that the thresholds $\gamma_A$ and $\gamma_B$ are not influenced by
the sequential nature of this task.  These thresholds affect the respective
probabilities of correct and wrong classifications but not the form of $\hat{d'}_{mt}$.

In Figure 6.3, the 90 per cent confidence intervals for these experimental
results are plotted on probability paper along with a fourth order polinomial
regression fit to Green's data (Green, 1960 a, McGill, 1968).  The predicted
performance for a sonar detection problem is also shown (Stallord and Leslie,
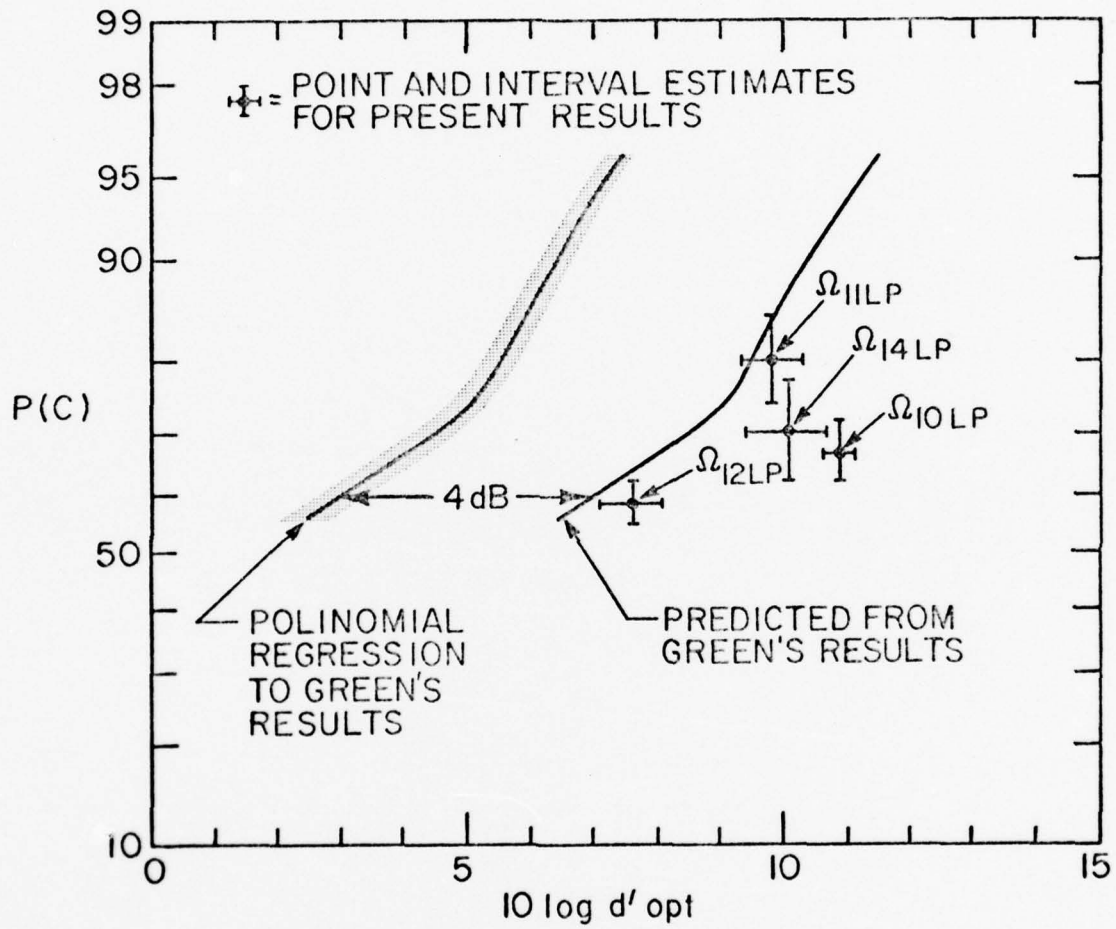1974.)

Figure 6.3   Comparison of High Pass Case Results with the
             Predicted Performance for the Detection of
             Noise in Noise

6.3  Analysis of Results Obtained in Experiments Using Shaped Noise

In the experiments with shaped noise vs. the recorded marine sound, the dichotomous feature was either the presence of amplitude modulation ($\Omega_{AM}$) in the case of signal pair $\omega_{10}$, $\omega_{10SN}$ or the presence of a strong tonal ($\Omega_{TONE}$) for signal pair $\omega_{14}$, $\omega_{14SN}$ . Neither one of these features is as simple as implied, however. The spectrum of these sources exhibits frequency as well as amplitude variations with time. Also, signal $\omega_{14}$ has a family of tones with only one pronounced steady tone and some varying components.

The feature $\Omega_{AM}$ appears as a repeated burst of noise impressed upon a continuous spectrum. Miller and Taylor (Miller, 1948, Miller and Taylor, 1948) have investigated the subjective character of this type of signal. It was found that the differential threshold for intensity $\Delta I/I$ increases as the duration of the added burst of noise decreases. In the limit where the duration of the added increment exceeds about 250 msec., the performance is given by the Weber fraction

$$\Delta I/I \cong 0.1 , \quad t > 250 \text{ msec.} \quad (\text{Green, 1960a}) .$$

For the signal in question, the natural modulation corresponding to feature $\Omega_{AM}$ is in the form of short bursts of noise which are repeated more or less periodically at a rate of 8 to 15 Hz. The duration of the noise pulse is approximately 25 msec. but with a non-rectangular waveform. From Figure 5.1 we see that the effective bandwidth of the signal vs. noise background is approximately 6 kHz from 400 Hz to 6400 Hz at the 3 db down points. The Weber fraction $\Delta I/I$ is then the incremental change in intensity vs. average intensity just perceptible. Moore and Raab (1975) find that the Weber fraction is:

-1.4 db for 3160 Hz bandwidth and duration 10 msec.
-0.9 db for 1000 Hz bandwidth and duration 10 msec.
-5.5 db for 3160 Hz bandwidth and duration 250 msec.
-4.9 db for 1000 Hz bandwidth and duration 250 msec.
-7.3 db for a 18 kHz bandwidth and duration 250 msec.

To compare the present results with these findings, we must know the amplitude excursion associated with the noise burst. The spectral level near the

center 12.5 msec time period of the amplitude modulation pulse was measured
to be 4 db higher than the average spectral level. From the results shown in
Table 5.4, the following can be calculated:

a. The average signal plus noise level was +1.65 db relative the back-
ground level alone. The action of the automatic loudness level control circuit
would be nearly to eliminate this effect.

b. The point estimate of instantaneous signal plus noise level was then
+3.5 db relative the background level or 1.85 db relative the average signal
plus noise ratio. This results in a observed Weber fraction of -2.7 db. The
90 per cent confidence interval on the observed Weber fraction in our case is:

$$-3.5 \text{ db} < W_{obs} < -2.6 \text{ db} .$$

When these results are compared to those of Moore and Raab (1975) it is
seen that the present observed SNR to respond agrees reasonably well. The
pulses comprising the modulation fall betwen these investigated by these authors
and therefore cannot be directly compared with those. However, extrapolating
their results using the empirical method proposed by them, the predicted Weber
fraction would be about -4.0 db. The difference between this predicted value
and the results obtained using the modified threshold procedure is certainly
not unreasonable considering the simplifying assumptions. Hence, as for the
classification of the high pass filtered signal, the results observed can be
explained reasonably well by simply assuming that the problem is one of de-
tecting a dichotomous feature. It must be noted, however, that treating these
pulse-like increases in the amplitude as isolated pulses to be detected in a
noise background must be done with caution. These pulses occur often enough
so that they are approaching an indistinguishable series of pulses as investigated
by Miller and Taylor (1948). The critical frequency is about 20 Hz and is also
a function of duty factor.

In the case of feature $\Omega_{TONE}$ , the tone at about 2.5 kHz has a spectral
level which is some 21 db per Hz greater than is the 1 Hz broadband spectral
level of the signal. The point estimate of the SNR required for classification
was -0.45 db. The calibration in this case was made in the 1/3rd octave band
which contained the tone. However, to obtain the average broadband background

against which the tone is heard, we take the average spectral level in the adjacent 1/3rd octave bands as representing the signal background level. When treated in this way, we obtain:

    a.  The signal plus noise to noise background at the time of classification responses was +1.4 db relative the background level.  This is the point of normalization.

    b.  The spectral level of the principal tone was then 19.5 db in a 1 Hz band relative the background or about 18 db relative the signal plus noise to noise reference point.

Hawkins and Stevens (1950) investigated the required spectral level of a tone for it to be just audible in a broad noise background.  They found that an average relative spectral level of about 20 db is required for a tone at 2.5 kHz to be audible.  The results found in the present experiment are consistent with these findings.  The difference noted is well within experimental errors in calibration and spectral measurement.

While the published data against which the present results are compared do not specifically define a detectability d' for the respective features, similar results have been obtained by workers using signal detectability theories (Green, 1968, Green, 1970, Swets, Green and Tanner, 1962, Tanner, 1961). The results obtained by Moore and Raab (1975) are somewhat at variance with a model of the auditory process as an energy detector and a difference of some 3 to 5 db is observed by them between their studies and the energy detection model.  In any case, the present results agree reasonably well with the findings of a number of workers who have concentrated on the signal detection task.

CHAPTER VII

SUMMARY AND CONCLUSIONS

## 7.1  Summary

In this report the complex task of aural classification of a source of
sound is reduced to a simpler problem of deciding between two sound patterns.
It was further assumed that in some cases this process could be treated
as a detection problem where the listener must first have the opportunity to
detect a signal feature before he can make a classification.  When the
classification is considered in this way, it was found that for six cases,
three entirely different dichotomous features, the observed classification
performance is within a few dB of results predicted by other researchers.
The studies against which the present experimental results are compared were
done with much simpler signals and all treated only the detection problem.

The good agreement between the present results and those obtained by
others argues in favor of an interpretation of these experiments as the
detection problem of a dichotomous feature.  Certainly this is an over-
simplification, because the criterion with which the listeners responded seems
to depend on other characteristics of the signal.  It is a simplification
which, however, is useful in the interpretation of a very complex problem
in cognition.

If indeed the results can be interpreted on the basis of this simplifying
model, the implications are far ranging.  It is possible then, for example,
to predict directly from the results of fundamental studies what the effect
of cutoff frequency of a filter will be on classification performance.  A
wealth of such results is available in the literature some of which is listed
in the bibliography of this report.  Likewise, the effect of certain types of
modulation, and of tones can be directly inferred.

Another significant finding to date is the fact that listeners are
equally proficient in making a decision about the absence of a feature as they
are about its presence.  This result was obtained for the case where the
two signals had equal a priori probability of occurrence, and the penalty for
each type of error was nominally the same.

## 7.2  Limitations on the Present Result

When attempting to relate the present findings to the anticipated performance of a sonar operator, some distinguishing features must be considered.  The listeners used here were trained college students with considerable experience in critical listening.  They have not been trained in the sonar classification task, however, and the signal patterns to which they were exposed were essentially context free.  Also, they did not need to classify on the basis of learned and retained information; rather it was a matching problem with, at most, intermediate duration memory.

A sonar operator generally would need to classify targets into one of a substantial set of possible targets although at times a two-point classification may be possible.  The sonar operator also is able to alternately listen to the target of interest and the ambient noise by means of a trainable beam (in most systems.)

The extent to which the training of a sonar operator allows him to outperform the college students will be investigated in experiments to be conducted at the Sonar School in San Diego in August 1975.  These tests and their principal objectives are described in an appendix hereto.

The experimental technique used, the modified threshold procedure, allows only a determination of listener performance at the time that a terminal decision is made.  While this is itself a useful bit of information, in many cases it would be desirable to know how the performance (i.e. $P(c)$) varies with signal-to-noise ratio.

## 7.3  Suggestions for Future Studies

To address the question of how the classification performance varies with signal-to-noise ratio, two additional sets of experiments are proposed.

The first of these will be conducted at the Sonar School in conjunction with the modified threshold procedure.  The equipment uses the same signal presentation sequence as is shown in Figure 4.1.  The listener will, however, continuously indicate his degree of confidence and his tentative classification decision by means of a linear potentiometer graduated as below.

```
       A            B
  10 8 6 4 2 0 2 4 6 8 10
```

The listener will indicate, say, a 2 toward the B if he has low confidence that the probe signal corresponds to exposure signal B and will move up the scale as his confidence increases. Some preliminary tests of this scheme ascertain that the listeners are able to change from an initial classification of say A to the alternate classification. Interpretation of these results is not trivial, however, since the confidence scale tends to measure the process:

> State a confidence of a classification given
> that an immediately preceding confidence was
> thought to apply.

The second test to obtain more knowledge about the psychometric function will be conducted at The Pennsylvania State University thanks to the efforts of Dr. James Martin. These will be two alternative forced response tests (2AFC) using a large number of naive college students. While the performance for such listeners is expected to be less predictable, the experiment will also be simpler. These results will principally test the degree to which the modified threshold procedure conforms to standard signal detectability measures.

An additional test will be made in the sonar school studies of the effect of shadowing alluded to in a previous chapter. An adequate number of events will be interspersed with other modified threshold tests to address this point. For these events, an additional 30 second delay will be introduced by a timer between the onset of the response period and the beginning of the step increase in SNR. It is then possible to test the effect of a prolonged shadowing period independently using the t test for difference of means.

# REFERENCES

Abramowitz, M., and I. A. Stegun, Handbook of Mathematical Functions, National Bureau of Standards, U.S. Department of Commerce, 1964.

Ahumada, A., R. Marken, and A. Sandusky, "Time and Frequency Analyses of Auditory Signal Detection," J. Acoust. Soc. Amer., 57:2, 1975, 385-390.

American Academy of Ophthamology and Otolarnygology, Committee on Conservation of Hearing, "Guide for the Conservation of Hearing in Noise," Los Angeles: Research Center, Subcommittee on Noise in Industry, 1957.

Andrews, F. A., and J. D. Hovater, Aural Detection by Passive Sonar, SAO Tech. Memo. 71-16, ASW Systems Project, Naval Material Command, Washington, D. C., 1971.

Annett, J., and L. Paterson, "Training for Auditory Detection," Acta Psychologica 27, 1967, 42-426.

Boehme, H., and F. Weidmann, "Probability of Detection and Probability of False Alarm for Passive Sonar Systems," JUA (USN), 20 (3), 1970, 573.

Boring, E. G., A History of Experimental Psychology, New York: Appleton-Century-Crofts, 1950.

Corso, J. F., The Experimental Psychology of Sensory Behavior, New York: Holt, Rinehart and Winston, 1967.

Cremer, L., P. Plenge, and D. Schwarzl, "Kurven Gleicher Lautstarke Mit Oktav-Gefiltertem Rauschen Im Diffusen Schallfeld," Acustica 9, 1959, 65-75.

Davis, H., and S. R. Silverman (Ed.), Hearing and Deafness, New York: Holt, Rinehart and Winston, 1961.

Durlach, N. I., "Note on Binaural Masking-Level Differences as a Function of the Interaural Correlation of the Masking Noise," J. Acoust. Soc. Amer. 1964, 1613-1617.

Eriksen, C. W., and H. W. Hake, "Multidimensional Stimulus Differences and Accuracy of Discrimination," J. Exp. Psychol., 50, 1955, 153-160.

Fletcher, H., "Auditory Patterns," Rev. Mod. Phys., 12, 1940, 47-65.

Fletcher, H., and W. A. Munson, "Relation Between Loudness and Masking," J. Acoust. Soc. Amer. 9, 1937, 1-10.

Fu, K. S., Sequential Methods in Pattern Recognition and Machine Learning, Academic Press, New York, 1968.

Goldstein, A. J., and J. R. Mundie, "Response of the Primary Auditory Neuron to Human Speech," Aerospace Medical Research Laboratory Report AMRL-TR-71-73, 1971.

Green, D. M., "Detection of Multiple Component Signals in Noise," J. Acoust. Soc. Amer., 30, 1958, 904-911.

a.  Green, D. M., "Auditory Detection of a Noise Signal," J. Acoust. Soc. Amer., 32, 1960, 121-131.

b.  Green, D. M., "Psychoacoustics and Detection Theory," J. Acoust. Soc. Amer., 32, 1960, 1189-1203.

Green, D. M., "Application of Detection Theory in Psychophysics," Proc. of the IEEE, 58, 1970, 713-723.

Green, D. M., and J. A. Swets, Signal Detection Theory and Psychophysics, New York: John Wiley and Sons, 1966.

Greenwood, D. D., "Auditory Masking and the Critical Band," J. Acoust. Soc. Amer., 33, 1961, 484-502.

Gundy, R. F., "Auditory Detection of an Unspecified Signal, J. Acoust. Soc. Amer., 33, 1961, 1008-1012.

Harris, J. D., Harris on Audiology, Groton Conn., The Journal of Auditory Research, 1973.

Hawkins, J. E., Jr., and S. S. Stevens, The Masking of Pure Tones and of Speech by White Noise," J. Acoust. Soc. Amer., 22, 1950, 6-13.

Helstrom, C. W., Statistical Theory of Signal Detection, New York:  Pergamon Press., 1960.

Hogg, R. V., and A. T. Craig, Introduction to Mathematical Statistics, London: Macmillan, 1970.

Julesz, B., "Problems of Pattern Recognition and Perceptual Psychology," in L. N. Kanal (Ed.) Pattern Recognition, Washington, D. C.:  Thompson Book Co., 1968.

Kaplan, H. M., Anatomy and Physiology of Speech, New York:  McGraw-Hill, 1960.

Lathi, B. P., Random Signals and Communication Theory, Scranton, PA:  International Textbook Co., 1968.

Leeper, R., "Cognitive Processes," in S. S. Stevens (Ed.), Handbook of Experimental Psychology, New York:  Wiley, 1951.

Licklider, J. C., "Basic Correlates of the Auditory Stimulus," in S. S. Stevens (Ed.), Handbook of Experimental Psychology, New York:  Wiley, 1951.

Licklider, J. C. R., D. Bindra, and I. Pollack, "The Intelligibility of Rectangular Speech-Waves," Amer. J. Psychol., 1948, 61, 1-20.

Loeb, M., and J. R. Binford, "Examination of Some Factors Influencing Performance on an Auditory Monitoring Task with One Signal Per Session," J. Exp. Psych. 83:1, 1970, 40-44.

Miller, G. A., "The Perception of Short Bursts of Noise," J. Acoust. Soc. Amer., 20:2, 1948, 160-170.

Miller, G.A., and W.G. Taylor, "The Perception of Repeated Bursts of Noise,"
J. Acoust. Soc. Amer., 20:2, 1948, 171-182.

McGill, W.J., "Polynomial Psychometric Functions in Audition," J. Mathemati-
cal Psych., 5, 1968, 369-376.

Moore, B.C.J., and D.H. Raab, "Intensity Discrimination for Noise Bursts,"
J. Acoust. Soc. Amer., 57:2, 1975, 401-406.

Murdock, B.B., and J.C. Ogilvie, "Binomial Variability in Short-Term Memory,"
Psych. Bull., 70:4, 1968, 256-260.

Myasnikov, L.L. and Ye. N. Myasnikova, Automatic Recognition of Sound Images,
Joint Publications Research Service Report JRS-54402, 1971, Translation of the
Russian language book Avtomaticheskoye Raspoznavaniye Zvukovykh, Leningrad: 1970.

National Research Council, Human Factors in Undersea Warfare, Washington, DC,
1949.

Parkinson, S.R., T.E. Parks, and N.E.A. Kroll, "Visual and Auditory Short-
Term Memory," J. Exp. Psych., 87:2, 1971, 274-280.

Peterson, A.P.G., and E.E. Gross, Handbook of Noise Measurement, West Concord,
Mass.: General Radio Co., 1963.

Peterson, W.W., T.G. Birdsall, and W.C. Fox, "The Theory of Signal Detectability,"
Trans. I.R.E., PGIT-4, 1954, 171-212.

Plomp, R., and G.F. Smoorenburg, (Ed.), Frequency Analysis and Periodicity Detection
in Hearing, Leiden: A.W. Sijthoff, 1970.

Pollack, I., "Effects of High Pass and Low Pass Filtering on the Intelligibility
of Speech in Noise," J. Acoust. Soc. Amer., 20:3, 1948, 259-266.

Pollack, I., "Identification of Elementary Auditory Displays and the Method of
Recognition Memory," J. Acoust. Soc. Amer., 31, 1959, 1126-1128.

Raab, D.H., and I.A. Goldberg, "Auditory Intensity Discrimination with Bursts
of Reproducible Noise," J. Acoust. Soc. Amer., 57:2, 1975, 437-447.

Raisbeck, G., Information Theory, An Introduction for Scientists and Engineers,
Cambridge, Mass.: MIT Press, 1972.

Robinson, D.E., and C.S. Watson, "Psychophysical Methods in Modern Psychoacoustics,"
in J.V. Tobias (Ed.), Foundations of Modern Auditory Theory, Vol. II, New York:
Academic Press, 1972.

Sholl, H.A., "Man-Machine Systems for Detection, Recognition, Transmission, and
Perception of Information," Vol. II, Part III, Connecticut University Report, 1971.

Siebert, W.M., "Frequency Discrimination in the Auditory System: Place or
Periodicity Mechanisms," Proc. of the IEEE, 58, 1970, 723-730.

Sorkin, R.D., "Extension of the Theory of Signal Detectability to Matching
Procedures in Psychoacoustics," J. Acoust. Soc. Amer., 34, 1962, 1745-1751.

Stallard, J. M. and C. B. Leslie, "Psychoacoustics and Passive Sonar Detection," Naval Ordnance Laboratory Technical Report NOLTR 74-27, 1974.

Stevens, K. N. and A. S. House, "Speech Perception," in J. V. Tobias (Ed.), Foundations of Modern Auditory Theory, Vol. II, New York: Academic Press, 1972.

Stevens, S. S., "Neural Events and Psychophysical Law," Science 170, 1970, 1043-1050.

Swets, J. A. (Ed.), Signal Detection and Recognition by Human Observers, New York: John Wiley, 1964.

Swets, J. A. and D. M. Green, "Sequential Observations by Human Observers of Signals in Noise," in C. Cherry, (Ed.), Information Theory, London: Butterworths, 1961.

Swets, J. A., D. M. Green, and W. P. Tanner, Jr., "On the Width of Critical Bands," Jr. Acoust. Soc. Amer., 34, 1962, 108-113.

Swets, J. A. and S. T. Sewall, "Stimulus vs. Response Uncertainty in Recognition," J. Acoust. Soc. Amer., 32, 1961, 1586-1592.

Swets, J. A., E. F. Shipley, M. J. Mikey, and D. M. Green, "Multiple Observations of Signals in Noise," J. Acoust. Soc. Amer., 31, 1959, 514-521.

Swets, J. A., W. P. Tanner, Jr., and T. G. Birdsall, "Decision Processes in Perception," Psychol. Rev. 68, 1961, 301-309.

Tanner, W. P., Jr., "Application of the Theory of Signal Detectability to Amplitude Discrimination," J. Acoust. Soc. Amer., 33, 1961, 1233-1244.

Tanner, W. P., Jr., "Theory of Signal Detectability as an Interpetive Tool of Psychophysical Data," J. Acoust. Soc. Amer., 32, 1960, 1140-1147.

Tanner, W. P. and R. D. Sorkin, "The Theory of Signal Detectability," in J. V. Tobias (Ed.), Foundations of Modern Auditory Theory, Vol. II, New York: Academic Press, 1972.

Tobias, J. V., (Ed.), Foundations of Modern Auditory Theory, New York: Academic Press, 1972.

Treisman, M. and F. R. Watts, "Relation Between Signal Detectability Theory and the Traditional Procedures for Measuring Sensory Thresholds," Psychol. Bull., 66:6, 1966, 438-454.

Urick, R. J. and G. C. Gaunaurd, "Detection of Fluctuating Sonar Targets," Naval Ordnance Laboratory Technical Report NOLTR 72-47, 1972.

Van Trees, H. L., Detection, Estimation, and Modulation Theory, Part I, New York: Wiley, 1968.

Wald, A., Sequential Analysis, New York: John Wiley, 1947.

Weinberg, G. H. and J. A. Schumaker, Statistics an Intuitive Approach, Benton, CA: Wadsworth Publishing Co., 1962.

Winer, J. B., Statistical Principles in Experimental Design, New York: McGraw-Hill, 1962.

Zhukov, S. and L. A. Christovitch, "Some New Concepts in Auditory Analysis of Acoustic Flow," Sov. Phys. Acoust., 20.3, 1974, 237-240.

Zwicker, E., G. Flottorp, and S. S. Stevens, "Critical Band Width in Loudness Summation," J. Acoust. Soc. Amer., 29, 1957, 548-557.

Zwicker, E., "Ueber Psychologische und Methodische Grundlagen der Lautheit," Acustica, 8:1, 1958, 237-258.

Zwislocki, J., "Temporal Summation of Loudness: An Analysis," J. Acoust. Soc. Amer., 46, 1969, 431-441.

APPENDIX A

SUMMARY OF MODIFIED THRESHOLD TESTS

AT THE SONAR SCHOOL

Plans are to conduct extensive tests of the concepts proposed in this report. These tests will be held in August 1975 at the ASW School, San Diego, California. Tentatively, it is planned to use 8 senior sonar operators for five days to obtain in excess of 800 events. Of these, 720 events will make use of the modified threshold procedure to determine the SNR required to make a terminal classification decision.

The tests are designed to test the following:

a) How do results obtained to date using University students compare with those using trained sonar operators?

b) Verify aspects of the classification model, as proposed, to identify the range of applicability in the sonar context.

c) Attempt to measure the significance of various features in the classification task.

d) Provide more data points for signals of special significance to the sponsor.

e) Attempt to determine the effect of memory and shadowing as related to this type of task.

The treatments to be applied to the population of trained sonar operators are summarized in Table A.3. Treatments 1-9 will be tested with a sufficient number of events to allow estimation of the respective probabilities $P(C)$ and $P(E)$, the probability of a correct terminal decision and of an incorrect terminal decision. Treatments 10-17 will be used specifically to test for differences in the mean signal-to-noise ratio $(\overline{SNR})$ required for a terminal decision. The t test for difference of means applies in these cases.

Feature
Ω                    Characterization of the Feature

| Feature Ω | Characterization of the Feature |
|---|---|
| 1 | Presence of an octave band stationary noise with $f_c$ = 500 Hz. |
| 2 | Presence of an octave band stationary noise with $f_c$ = 4k Hz. |
| 3 | Square wave 10 Hz. amplitude modulation of an octave band with $f_c$ = 500 Hz. |
| 4 | Square wave 10 Hz. amplitude modulation of an octave band with $f_c$ = 4k Hz. |
| 5 | Presence of an octave band stationary noise with $f_c$ = 250 Hz. |
| 6 | Blade rate amplitude modulation of an octave band with $f_c$ = 4k Hz. |
| 7 | Marine source spectrum without amplitude modulation |
| 8 | Blade rate amplitude modulation of a broadband marine source |
| 9 | Marine source high pass spectrum with $f_\sim$ = 707 Hz. and amplitude modulation |
| 10 | Marine source bandpass spectrum with $f_c$ = 500 Hz. and amplitude modulation |
| 11 | Marine source bandpass spectrum with $f_c$ = 250 Hz. and amplitude modulation |
| 12 | Marine source low pass spectrum with $f_h$ = 177 Hz. and amplitude modulation |

Table A.1.   Summary of Dichotomous Features

Pattern                    Features Comprising the Pattern

| Pattern ω | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 1 | | | | | | | | | | |
| 2 | | 1 | | | | | | | | | | |
| 3 | 1 | 1 | 1 | | | | | | | | | |
| 4 | 1 | 1 | | 1 | | | | | | | | |
| 5 | | 1 | | 1 | | | | | | | | |
| 6* | .25 | 1 | | | | | | | | | | |
| 7 | | 1 | | | 1 | | | | | | | |
| 8 | 1 | 1 | | | | 1 | | | | | | |
| 9 | | 1 | | | | 1 | | | | | | |
| 10 | | | | | | | 1 | 1 | | | | |
| 11 | | | | | | | 1 | | | | | |
| 12 | | | | | | | | | 1 | | | |
| 13 | | | | | | | | | 1 | | 1 | 1 |
| 14 | | | | | | | | | 1 | 1 | | |
| 15 | | | | | | | | | 1 | 1 | 1 | 1 |

*A feature is present (1) or has 1/4th relative detectability (.25).

Table A.2.   Summary of Signal Patterns

| TREATMENT | PURPOSE | SIGNAL $\omega:H_o$ | PATTERNS $\omega:H_1$ | SIGNAL TO NOISE RATIO RANGE (dB) | NO. OF EVENTS |
|---|---|---|---|---|---|
| 1 | To test the effect of the presence of a band of noise | 2 | 1 | -5.0, 8.0 | 72 |
| 2 | To test the effect of amplitude modulation on a band of noise | 1 | 3 | -8.0, 5.0 | 72 |
| 3 | To test the effect of the presence of a band of noise in the presence of amplitude modulation | 5 | 4 | -5.0, 8.0 | 72 |
| 4 | To test the effect of the presence of an amplitude modulated band of noise | 2 | 3 | -8.0, 5.0 | 72 |
| 5 | To test the effect of changing the detectability of a dichotomous feature | 2 | 6 | -2.0, 11.0 | 60 |
| 6 | To determine the effect of actual modulation waveforms applied to a band of noise | 9 | 8.11+ | -5.0, 8.0 | 82 |
| 7 | To determine the effect of stripping the modulation from a recorded signal | 11.11 | 10.11 | -9.0, 4.0 | 72 |
| 8 | To evaluate the performance for recorded signals with a missing band of frequencies | 13.12 | 15.12 | -2.0, 11.0 | 72 |
| 9 | To determine how the criterion differs between University students and sonar operators | 9.12 | 15.12 | -2.0, 11.0 | 72 |
| 10 | To test for the difference in SNR between University students and sonar operators, high pass case | 9.11 | 15.11 | -3.0, 10.0 | 12 |
| 11 | To test for the difference in SNR between University students and sonar operators, high pass case | 9.10 | 15.10 | -4.0, 9.0 | 12 |
| 12 | To test for the difference in SNR between University students and sonar operators, shaped noise case | 11.10 | 10.10 | -9.0, 4.0 | 12 |
| 13 | To determine the effect of an extended shadowing period on performance | 2 | 6 | -2.0, 11.0 | 12 |
| 14 | To test the effect of the presence of a band of noise centered at 250 Hz | 2 | 7 | -5.0, 8.0 | 16 |
| 15 | To determine the effect of an extended shadowing period on performance | 9 | 8.11 | -5.0, 8.0 | 12 |
| 16 | To evaluate effect on SNR of changing the high pass cutoff frequency | 14.11 | 15.11 | -2.0, 11.0 | 24 |
| 17 | To evaluate effect on SNR of changing the high pass cutoff frequency | 14.12 | 15.12 | 0.0, 13.0 | 24 |

+ Features which use recorded marine sounds are identified by the corresponding signal identification

Table A.3.  Summary of Treatments Using the Modified Threshold Procedure Scheduled for Sonar School Tests

## APPENDIX B

## HARDWARE SYSTEMS

The tapes used as the source of the auditory stimulii in these tests are the end product of a carefully controlled series of taping steps. The marine sources of interest come recorded on a large variety of machines, at various tape speeds and recording techniques. Selected portions of these raw source tapes are re-recorded onto 1/2 inch magnetic tapes in FM recording mode and at 60 or 30 inches per second. During the re-recording of these master tapes, adjustment is made for hydrophone or other frequency weighting appearing on the raw data. The signals are also prewhitened by an amplifier with a 6 dB per octave increasing gain vs frequency behavior.

The master tapes serve as inputs to any stimulii recordings which use actual recorded marine sounds. Alternately, a weighted noise source may be substituted for either member of the exposure set. Recorded ocean ambient or shaped noise serves as the background noise against which the probe signal is presented. These various signals are properly time multiplexed by means of an analog selector which receives control signals from a digital sequencer. Figure A.1 shows the various components of the system needed to create primary tapes from master tapes. The flow of analog and digital signals is also shown. In Figure A.1, those components which are not specifically indicated as being commercial items were fabricated at the Applied Research Laboratory. Most of these components were designed and built by the author. The major portions of the system with the exception of the analog input tape unit are pictured in Figure A.2. Necessary interconnecting cabling has been omitted from this photograph, however. Figure A.3 shows the sequencer/controller in greater detail.

The digital sequencer and associated controller orchestrates the required functions in response to a simple program stored in a read-only memory. The sequencer starts and stops the output tape drive, sequences the various audio signals, and causes the generation of control signals which will be used at the test site to indicate response periods and to record responses. The exact order of the process is determined by a program selector and by the state of sense switches. The setting of the balanced mixer is performed manually in
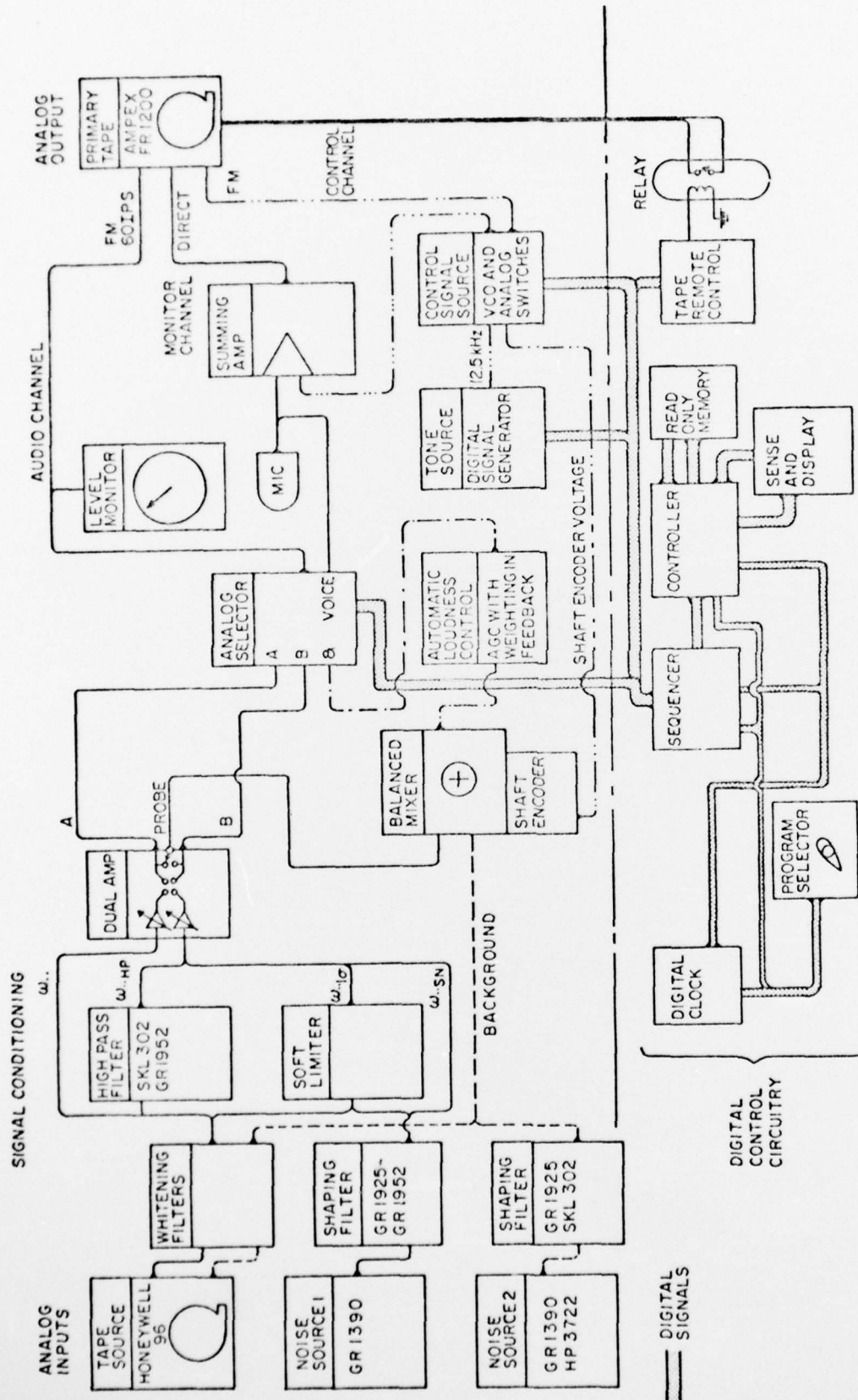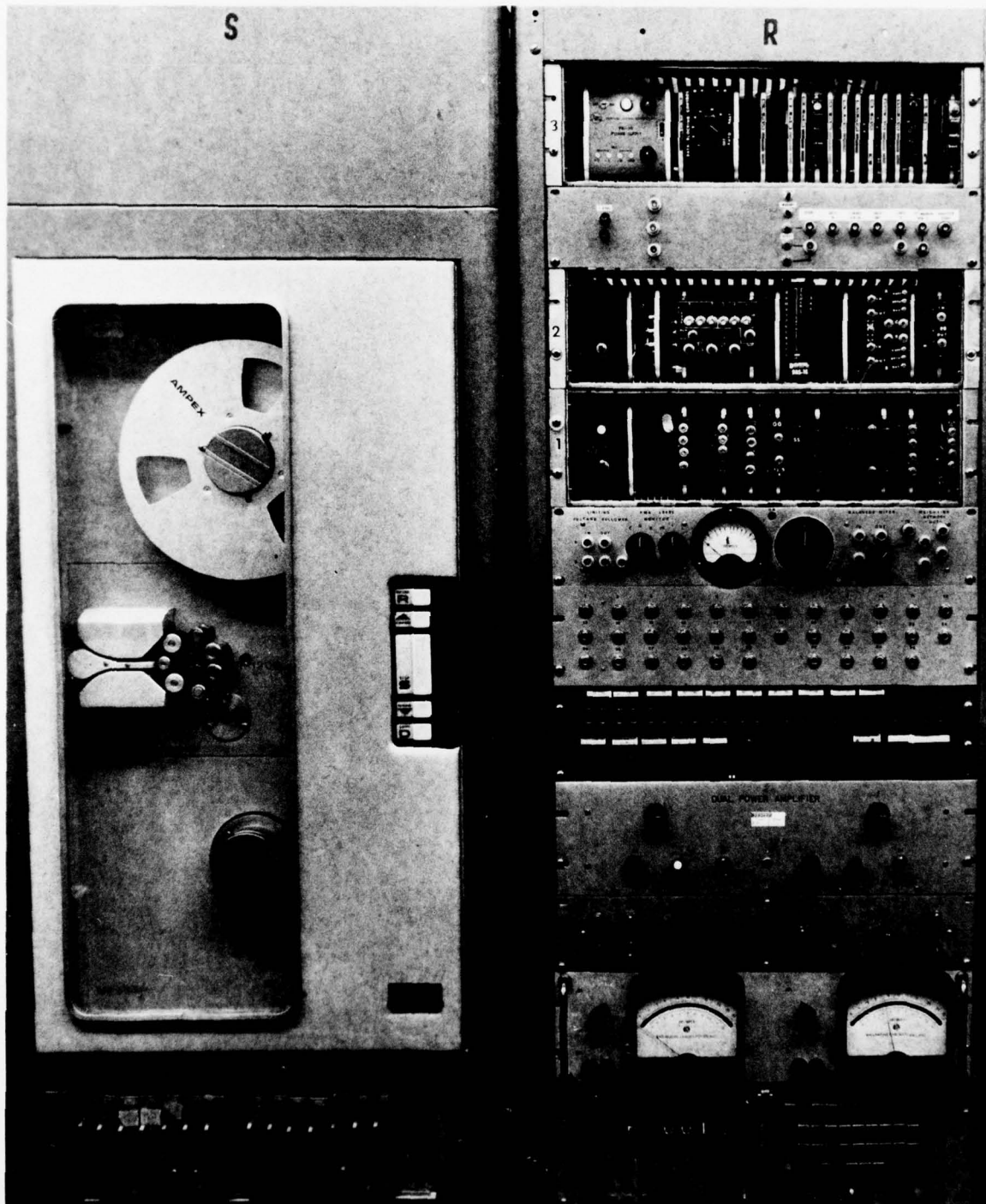
Figure A.1   Flow Diagram, Creating Primary Tapes

A.2  Photograph of the Hardware
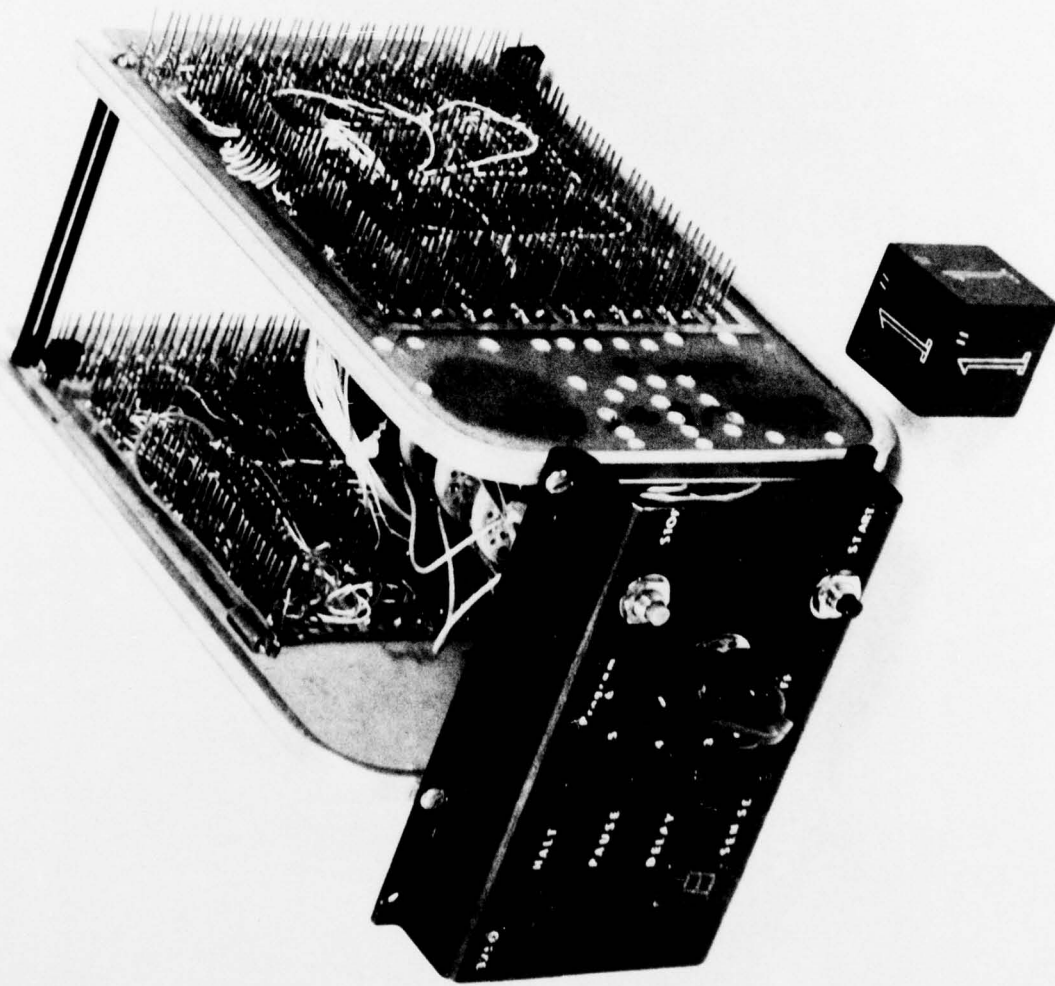
Figure A.3  Photograph of the Sequencer

response to a que light which indicates the response period portion of the audio sequence.

A direct recorded monitor channel of the output tape is used for voice annotation of cut numbers on the primary tapes and to record a 12.5 kHz pilot tone. This tone is on during valid portions of data on an FM recorded audio channel and an FM recorded control channel. This control channel has tones for control of the response recorder and a balanced mixer setting proportional VCO output.

Cuts from the primary tapes are re-recorded onto the final audio tapes in a randomized order. The monitor channel of the primary tape is used both to locate the desired cuts and to allow for voice annotation between events on the audio tapes. This voice annotation of event numbers and of special instructions to the listener is usually from a cassette recording. A phase locked loop, analog selector combination, insures that the objectionable FM discriminator noise output while searching the primary tape does not appear on the tapes used for listener tests. Figure A.4 diagrams this aspect of the recording process.
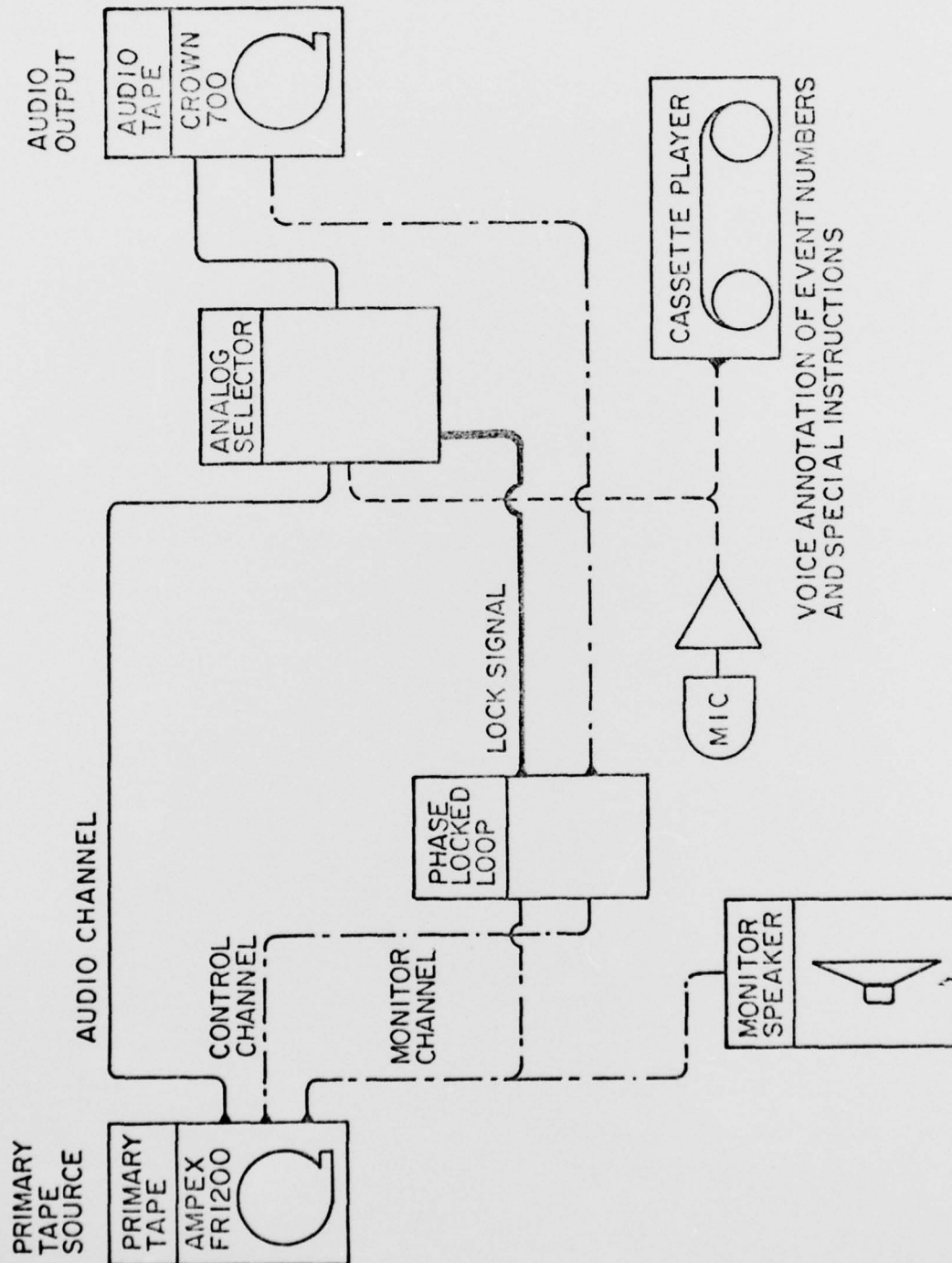
Figure A.4   Flow Diagram, Creating Final Audio Tapes

DISTRIBUTION   LIST    TM 75-165

CNO, Code NOP098, Copy No. 1
CNM, Code ASW-00, Copy No. 2
DDC, Copy No.3 and 4
NAVAIR, Mr. P. W. Quinn, AIR 370F, Copy No. 5
NAVSEA, Mr. Andre LaPointe, SEA 660E3, Copy No. 6
NAVSEA, Library, SEA 09G32, Copies No. 7 and 8
Third Fleet, Analysis Division, Code N7, Copy No. 9
ONR, Mr. J. G. Smith, Code 462, Copy No. 10
ONR, Dr. M. Tolcott, Copy No. 11
CNA, Dr. S. K. Shear, OP-03EG, Copy No. 12
Fleet Antisubmarine School, Mr. P. Assadorian, Copy No. 13
NADC, Mr. R. G. Houser, Code 20623, Copy No. 14
NADC, Library, Copy No. 15
NUSC, Mr. H. Wilms, Code PA11, Copy No. 16
NUC, Mr. R. Acks, Copy No. 17
Stanford Research Institute, Mr. J. R. Olmstead, Copy No. 18
NSWC, Mr. J. M. Stallard, Copy No. 19